8-2019

The chemical mechanisms of plant-soil interactions

Liam O'Connor Mueller
University of Tennessee, lmuelle3@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Liam O'Connor Mueller entitled "The chemical mechanisms of plant-soil interactions." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Ecology and Evolutionary Biology.

Jennifer A. Schweitzer, Major Professor

We have read this dissertation and recommend its acceptance:

Joseph K. Bailey, Brian C. O'Meara, Shawn R. Campagna

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

The chemical mechanisms of plant-soil interactions

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Liam O'Connor Mueller
August 2019

DEDICATION

To the memory of my mother
Maura E. O'Connor

ʻO ʻoe ka pueo o Pālolo
Ka mea hoʻokipa o Kamakou
He makana na nā kūpuna

ACKNOWLEDGEMENTS

I would like to express my deep thanks to my advisor, Jennifer Schweitzer, whose guidance made this possible. I would also like to acknowledge my committee, Joseph Bailey, Brian O'Meara, and Shawn Campagna, who provided diverse and excellent insight. This work also represents help from an incredible lab: Michael van Nuland, Rachel Wooliver, Alix Phenningwreth, Ian Ware, Lauren Breza, Mark Genung, Courtney Gorman, Shannon Bayliss, Kendall Beals, Greg Newman, Caroline Daws and Peter Meidl. I consider you all great friends. For their help in the greenhouse and in the field I must thank, Terrell Carter, Michaela Humby, Philip Patterson, Ken McFarland, Jeff Martin, Chris Wong, Ashley Shaw, Alysia Kaha, Riley DeMattos, James Akau, Kainana Francisco, and the 2014 HYCC Hilo3 team. Thanks to Charles Mueller for 'the little things', and Melissa Liotta for keeping me sane.

This work could not have happened without funding from the University of Tennessee, Ecology and Evolutionary Biology department, and from the USDA Forest Service Pacific Southwest Research Station, the Americorp Program managed by KUPU, Hawai`i, the Cowden Endowment at Northern Arizona University, National Science Foundation grant DEB–1020412 (T. Fukami, C. P. Giardina, G. P. Asner), and the Hooper Undergraduate Research Award.

ABSTRACT

The interaction between plants and soils is too often oversimplified for its importance to life on earth. The chemical complexity of this interaction is enormous, but many soil scientists, ecologists, and biogeochemists reduce it to single compounds. This dissertation looks to explore the complexity of the interface of plant roots and soils, termed the rhizosphere, in an effort to better understand the chemical forces that shape terrestrial ecosystems. In two chapters I explore how small genetic differences in plants can lead to vast differences in surrounding soil enzymes and thousands of other metabolites. I show that the trait variation within a plant species can alter pedogenesis (i.e., the formation of soils) that may have feedbacks on future generations of plants. In the third chapter I explicitly explore this feedback by examining how plant chemical phenotypes change when exposed to live soils which have been conditioned by populations of a plant with unique phenotypes. As soil metabolomics is a new field, I used my final chapter as a review of the current methods of data analysis with the hope that future soil scientists will fully explore the chemical complexity of the soil instead of ignore it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Understanding the connections between plants and soils is obscured by the difficulty of making observations belowground and the complex biological and chemical interactions of the root-soil interface. These challenges are worth overcoming because plant-soil interactions in part determine the composition and function of soil biotic communities (Schweitzer et al. 2008, Orwin et al. 2010) that mediate a broad range of ecosystem services including pedogenesis (Mueller et al. 2017), carbon and nitrogen cycles (Wooliver et al. 2018) and can determine plant performance and evolution (Lau and Lennon 2012, Ware et al. 2019). My dissertation seeks to improve our understanding of plant-soil interactions with exploration of the chemical consequences of plant-soil interactions, novel inquiry into the plant and soil metabolome, and the introduction and implementation of statistical methods new to the metabolomics and metagenomics fields. Our understanding of plant-soil interactions are from studies either ignoring the small scale complexity and mechanism in favor of landscape level patterns, or the opposite, with trees growing in sterile conditions, exploring only a single factor of the rhizosphere (See recent reviews in van Dam and Bouwmeester 2016, Ryan et al. 2016, Erktan et al. 2018). Luckily, sufficient work has occurred at both of these scales to allow studies that bridge this gap in our understanding. My dissertation aims to explore how small-scale belowground changes are shaped by landscape-level processes, and in turn, how landscapes can be shaped by small belowground chemical changes.

Plant conditioning of soil is essential to understand because of its ecological and evolutionary consequences for plant populations, community dynamics, and ecosystem function. Plant-soil linkages and feedbacks are important processes that can drive plant ranges and distributions (Bezemer et al., 2013; Van Nuland et al., 2017), determine success of invasive plant species (Klironomos 2002, Van der Putten 2003, Wolfe et al. 2008), and hasten plant community succession (Kardol et al. 2006). These processes are mediated by both above- and belowground traits that structure microbial communities and influence ecosystem dynamics (Vitousek et al. 1987, Whitham et al. 2006, Bardgett and Wardle 2010). For example, aboveground traits, such as leaf lignin concentrations, can influence belowground nutrient cycles by altering decomposition rates and mineralization of organic matter (Melillo et al. 1982) including within species (Hobbie et al. 2006). Plant phenotype belowground can also fundamentally alter soils (Genung et al. 2013) but the difficulty in studying and interpreting belowground phenotypes

makes this a large unknown variable in plant-soil studies. For example, plants can exude different mixtures of sugars and amino acids that can alter microbial communities and their interactions which mediates many ecosystem processes (Badri and Vivanco 2009, Rasmann and Turlings 2016, Zhalnina et al. 2018). However, it is only recently that we have begun to comprehensively explore the complex chemical interactions in the rhizosphere (van Dam and Bouwmeester 2016). As more studies look belowground, we are finding that the context-dependency of plant-soil interactions make it difficult to predict broad scale changes (Erktan et al. 2018). Exploring the mechanisms of population or individual level context-dependency will alleviate that difficulty. As Ryan and co-authors (2016) state: "The challenge is not only to understand how roots function but to do so in soil with all its physical, chemical and biological complexity." Doing this will allow us to scale up and explicitly include that context-dependency in our models of ecosystem function.

Plants have been shown to produce a wide variety of plant exudates that change the rhizosphere. Root exudate concentrations of sugars like fructose, glucose and maltose can change the rates of bacterial colonization (Lugtenberg et al. 1999). Other secondary metabolites released from plant roots, like organic acids, have been demonstrated to alter the soil characteristics around them (Zhalnina et al., 2018). Other plant root secondary metabolites associated with defence and allelopathy change among populations and over an individuals lifetime (Badri and Vivanco 2009, Dennis et al. 2010, van Dam and Bouwmeester 2016). Often it is changes in the soil itself that trigger these responses from plants. For example the presence of root herbivores altered root tissue exudates to contain a greater proportion of anti-herbivory compounds (Bezemer and van Dam 2005). To explore the range of molecules produced in the plant-soil interface and the consequences of this variation, my dissertation is split into four related chapters:

Chapter 1 was published in Ecosphere in 2017. We used a series of 3000-year-old, lava-created forest fragments on the Island of Hawai`i to examine whether disturbance and subsequent colonization can lead to genetically differentiated populations, and where differentiation occurs, if there are belowground consequences of trait driven changes. These fragments are dominated by a single tree species, *Metrosideros polymorpha* (Myrtaceae) or ʻōhiʻa, which have been actively colonizing the surrounding lava flow created in 1858. To test our ideas about differentiation of genetically determined traits, we: 1) created rooted cuttings of

ʻōhiʻa individuals sampled from fragment interiors and open lava sites, grew these individuals in a greenhouse, and then used these cuttings to create a common garden where plant growth was monitored for three years; and 2) assessed genetic variation and made Q_{ST}/F_{ST} comparisons using microsatellite repeat markers. Results from the greenhouse showed quantitative trait divergence in plant height and pubescence across plants sampled from fragment interior and matrix sites. Results from the subsequent common garden study confirmed that the matrix environment can select for individuals with 9.1% less shoot production and 17.3% higher leaf pubescence. We found no difference in molecular genetic variation indicating gene flow among the populations. The strongest Q_{ST} level (0.34 for height in the common garden) was far greater than the F_{ST} estimate (0.006), indicating sympatric genetic divergence in growth traits. Tree height was correlated with ecosystem properties such as soil carbon and nitrogen storage, soil carbon turnover rates, and soil phosphatase activity, indicating that selection for growth traits will influence structure, function, and dynamics of developing ecosystems.

Chapter 2 explores the chemical complexity of the root and rhizosphere metabolome across plant populations to determine evolutionary alterations of plant metabolism and subsequent linkage to soils. Metabolomics were used in conjunction with tree clones grown in a common environment to explore the root chemical phenotypes of six distinct populations of *Populus angustifolia*. By exploring the entire metabolome of the root instead of specific compounds, this study identified the magnitude of the chemical complexity present in roots with less sampling bias. This study will be the first of our knowledge to use metabolomics to examine variation in both roots and the soils across natural gradients when grown under common conditions. We found that genetically distinct populations of *P. angustifolia* have unique root phenotypes and population of origin could explain ~30% of the variation in the root holobiont metabolome (i.e., the combined root along with all microbial endophytes). Furthermore, because all tree clones were grown in a generic potting mix, we were able to show plant conditioning of rhizosphere soils. Individual plant root metabolome variation explained ~15% of the variation in the paired rhizosphere soil metabolome. This study showed population level genetic variation in root holobiont chemical phenotype is conditioning unique soil chemical phenotypes.

Where Chapter 1 showed how evolution can lead to unique pedogenesis and in Chapter 2 how local adaptation can lead to unique plant traits that can change their surroundings, in **Chapter 3** I explored how these conditioned soils influence plants and change their chemical

responses, through a reciprocal plant-soil interaction experiment. Using four populations of *P. angustifolia* grown in a greenhouse, I examined how the root tissue and rhizosphere soil metabolites change based on plant population and soil origin. I proposed two hypotheses: 1) the root tissue metabolome is a function of the interaction between the plant population and the soil environment it is exposed to; and 2) Soil rhizosphere metabolomes vary by both the soil microbial environment and the effect of tree root metabolome. A significant interaction term supported the first hypothesis that the root metabolome was influenced by both plant population and soil environment.

Finally, **Chapter 4** outlines the broad range and application of possible statistical analysis techniques appropriate for the analysis of metabolomics data sets. I review the current state of statistical analysis of large multivariate chemical datasets that can contain >10,000 chemical compounds. Metabolomics research published to date often simplify analyses by only analyzing a subset of the known compounds using analysis of variance (ANOVA), or use multivariate techniques like Partial Least Squares (PLS-DA) which may not adequately explore the data due to an inability to properly account for multiple groupings and nested levels of structure. Using data from chapter 2 as an example, I explored the consequences of multiple statistical and visual approaches by analyzing the same datasets with multiple approaches (e.g., NMDS, PCA, van Krevelen, PLS-DA, RDA, Mantel, co-inertia), focusing on how the statistical hypotheses change based on the analysis that is used, thus leading to different interpretations and utilizations of different components of the data. The goal of this review was to highlight the range of statistical techniques available to the rapidly growing field of metabolomics and provide guidance and 'best practices' for choosing question-specific appropriate analyses.

These four chapters aim to explore the degree to which local adaptation leads to changes in plant traits that feedback on the rhizosphere chemical community. I am addressing big topics in ecosystem ecology such as the evolutionary potential of plant soil feedbacks (Chapter 2 and 3), the scale of plant trait divergence across environmental gradients (Chapters 1, 2, and 3), and the importance of complex belowground conditions (Chapters 2 and 3), using novel analytical and soil techniques to capture the fine scale complexity of the rhizosphere (Chapters 2, 3, and 4). My dissertation applies a range of mechanistic methods and approaches to tackle some of ecosystem and chemical ecology's most pressing questions, while also advancing metabolomics

and metagenomics research by expanding the scope and relevance of possible questions asked and analyses utilized.

References

Badri, D. V., and J. M. Vivanco. 2009. Regulation and function of root exudates. Plant, Cell & Environment 32:666–681.

Bardgett, R. D., and D. A. Wardle. 2010. Aboveground-Belowground Linkages Biotic Interactions, Ecosystem Processes, and Global Change. Oxford University Press, Oxford.

Bezemer, T. M., and N. M. van Dam. 2005. Linking aboveground and belowground interactions via induced plant defenses. Trends in Ecology & Evolution 20:617–624.

van Dam, N. M., and H. J. Bouwmeester. 2016. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. Trends in Plant Science 21:256-265

Dennis, P. G., A. J. Miller, and P. R. Hirsch. 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? FEMS Microbiology Ecology 72:313–327.

Erktan, A., M. L. McCormack, and C. Roumet. 2018. Frontiers in root ecology: recent advances and future challenges. Plant and Soil 424:1–9.

Genung, M. A., J. K. Bailey, and J. A. Schweitzer. 2013. The Afterlife of Interspecific Indirect Genetic Effects: Genotype Interactions Alter Litter Quality with Consequences for Decomposition and Nutrient Dynamics. PLoS ONE 8:e53718.

Hobbie, S. E., P. B. Reich, J. Oleksyn, M. Ogdahl, R. Zytkowiak, C. Hale, and P. Karolewski. 2006. Tree species effects on decomposition and forest floor dynamics in a common garden. Ecology 87:2288–2297.

Kardol, P., T. Martijn Bezemer, and W. H. Van Der Putten. 2006. Temporal variation in plant-soil feedback controls succession. Ecology Letters 9:1080–1088.

Klironomos, J. N. 2002. Feedback with soil biota contributes to plant rarity and invasiveness in communities. Nature 417:67–70.

Lau, J. A., and J. T. Lennon. 2012. Rapid responses of soil microorganisms improve plant fitness in novel environments. Proceedings of the National Academy of Sciences 109:14058–14062.

Lugtenberg, B. J. J., L. V. Kravchenko, and M. Simons. 1999. Tomato seed and root exudate sugars: composition, utilization by Pseudomonas biocontrol strains and role in rhizosphere

colonization. Environmental Microbiology 1:439–446.

Martijn Bezemer, T., W. H. van der Putten, H. Martens, T. F. J. van de Voorde, P. P. J. Mulder, and O. Kostenko. 2013. Above- and below-ground herbivory effects on below-ground plant-fungus interactions and plant-soil feedback responses. Journal of Ecology 101:325–333.

Melillo, J. M., J. D. Aber, and J. F. Muratore. 1982. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. Ecology 63:621–626.

Mueller, L. O., L. C. Breza, M. A. Genung, C. P. Giardina, N. E. Stone, L. C. Sidak-Loftis, J. D. Busch, D. M. Wagner, J. K. Bailey, and J. A. Schweitzer. 2017. Ecosystem consequences of plant genetic divergence with colonization of new habitat. Ecosphere 8:e01743.

Van Nuland, M. E., J. K. Bailey, and J. A. Schweitzer. 2017. Divergent plant–soil feedbacks could alter future elevation ranges and ecosystem dynamics. Nature Ecology & Evolution 1:0150.

Orwin, K. H., S. M. Buckland, D. Johnson, B. L. Turner, S. Smart, S. Oakley, and R. D. Bardgett. 2010. Linkages of plant traits to soil properties and the functioning of temperate grassland. Journal of Ecology 98:1074–1083.

Van der Putten, W. H. 2003. Plant defense belowground and spatiotemporal processes in natural vegetation. Ecology 84:2269–2280.

Rasmann, S., and T. C. J. Turlings. 2016. Root signals that mediate mutualistic interactions in the rhizosphere. Current Opinion in Plant Biology 32:62–68.

Ryan, P. R., E. Delhaize, M. Watt, and A. E. Richardson. 2016. Plant roots: understanding structure and function in an ocean of complexity. Annals of Botany 118:555–559.

Schweitzer, J. A., J. K. Bailey, D. G. Fischer, C. J. LeRoy, E. V. Lonsdorf, T. G. Whitham, and S. C. Hart. 2008. Plant–soil–microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms. Ecology 89:773–781.

Vitousek, P. M., L. R. Walker, L. D. Whiteaker, D. Mueller-Dombois, and P. A. Matson. 1987. Biological Invasion by Myrica faya Alters Ecosystem Development in Hawaii. Science 238:802–804.

Ware IM, Van Nuland ME, Schweitzer JA, et al (2019) Climate-driven reduction of genetic variation in plant phenology alters soil communities and nutrient pools. Global Change Biology 25:1514-1528

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. LeRoy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, B. M. Potts, D. G. Fischer, C. A. Gehring, R. L. Lindroth, J. C. Marks, S. C. Hart, G. M. Wimp, and S. C. Wooley. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics 7:510–523.

Wolfe, B. E., V. L. Rodgers, K. A. Stinson, and A. Pringle. 2008. The invasive plant Alliaria petiolata (garlic mustard) inhibits ectomycorrhizal fungi in its introduced range. Journal of Ecology 96:777–783.

Wooliver, R. C., J. K. Senior, B. M. Potts, M. E. Van Nuland, J. K. Bailey, and J. A. Schweitzer. 2018. Soil fungi underlie a phylogenetic pattern in plant growth responses to nitrogen enrichment. Journal of Ecology 106:2161–2175.

Zhalnina, K., K. B. Louie, Z. Hao, N. Mansoori, U. N. da Rocha, S. Shi, H. Cho, U. Karaoz, D. Loqué, B. P. Bowen, M. K. Firestone, T. R. Northen, and E. L. Brodie. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nature Microbiology 3:470–480.

CHAPTER I

ECOSYSTEM CONSEQUENCES OF PLANT GENETIC DIVERGENCE WITH COLONIZATION OF NEW HABITAT

A version of this chapter was originally published by Liam O. Mueller, Lauren C. Breza, Mark A. Genung, Christian P. Giardina, Nathan E. Stone, Lindsay C. Sidak-Loftis, Joseph D. Busch, David M. Wagner, Joseph K. Bailey, and Jennifer A. Schweitzer:

Liam O. Mueller, Lauren C. Breza, Mark A. Genung, Christian P. Giardina, Nathan E. Stone, Lindsay C. Sidak-Loftis, Joseph D. Busch, David M. Wagner, Joseph K. Bailey, Jennifer A. Schweitzer (2017). Ecosystem consequences of plant genetic divergence with colonization of new habitat. Ecosphere, 8:5, 1-16.

LOM, JKB, and JAS developed the original idea. LOM, LCB, MAG, CPG, JKB, and JAS collected field and greenhouse data. NES, LCS-L, JDB, and DMW collected the genetic data. Statistical analysis was done by LOM, LCB, MAG, JKB, and JAS. Each author contributed the writing or editing of the manuscript. The only modification to this manuscript made for the dissertation is the APPENDIX where each figure can be found.

Abstract

When plants colonize new habitats altered by natural or anthropogenic disturbances, those individuals may encounter biotic and abiotic conditions novel to the species, which can cause plant functional trait divergence. Over time, site driven adaptation can give rise to population-level genetic variation, with consequences for plant community dynamics and ecosystem processes. We used a series of 3000-year-old, lava-created forest fragments on the Island of Hawai`i to examine whether disturbance and subsequent colonization can lead to genetically differentiated populations, and where differentiation occurs, if there are ecosystem consequences of trait driven changes. These fragments are dominated by a single tree species, *Metrosideros polymorpha* (Myrtaceae) or ʻōhiʻa, which have been actively colonizing the surrounding lava flow created in 1858. To test our ideas about differentiation of genetically determined traits, we: 1) created rooted cuttings of ʻōhiʻa individuals sampled from fragment interiors and open lava sites, raised these individuals in a greenhouse, and then used these cuttings to create a common garden where plant growth was monitored for three years; and 2) assessed genetic variation and made Q_{ST}/F_{ST} comparisons using microsatellite repeat markers.

Results from the greenhouse showed quantitative trait divergence in plant height and pubescence across plants sampled from fragment interior and matrix sites. Results from the subsequent common garden study confirmed that the matrix environment can select for individuals with 9.1% less shoot production and 17.3% higher leaf pubescence. We found no difference in molecular genetic variation indicating gene flow among the populations. The strongest Q_{ST} level (0.34 for height in the common garden) was far greater than the F_{ST} estimate (0.006), indicating sympatric genetic divergence in growth traits. Tree height was correlated with ecosystem properties such as soil carbon and nitrogen storage, soil carbon turnover rates, and soil phosphatase activity indicating that selection for growth traits will influence structure, function, and dynamics of developing ecosystems. These data show that divergence can occur on centennial time scales of early colonization.

Introduction

The movement of plant species into new habitats can have important evolutionary and ecological consequences. For example, when species colonize new created habitats, as a result of natural or anthropogenic disturbances, they may encounter novel biotic and abiotic factors that can: (*i*) affect their success or failure (Hobbs et al. 2006); (*ii*) influence the evolutionary dynamics of populations (Holt 2003, Parmesan 2006); and (*iii*), alter functional phenotypes (Whitham et al. 2006, Read et al. 2014, Bailey et al. 2014, Schweitzer et al. 2014, Kinnison et al. 2015). To better predict the consequences of colonization of new habitats from the influences of natural disturbance and global environmental change, understanding how selection operates during colonization is critical to correctly interpreting ecosystem consequences of phenotypic divergence.

The evolution of colonizing plant species has been examined in the context of local adaptation (Clausen, Keck and Hiesey 1947), coevolution (Carroll et al. 2005), and in the context of exotic species invasions (Zenni et al. 2014a). It has become apparent that individuals colonizing unique ecotypes are phenotypically different from those occurring in source populations within the previous core range of the species (Phillips et al. 2006, Eckert et al. 2008, Felker Quinn et al. 2013). For example, invasive pine trees in Brazil had distinct phenotypes in colonized versus source portions of their range (Zenni et al. 2014b.), suggesting that selective pressures during colonization favored certain traits. Such effects clearly have important implications for managing invasive species, but beyond the study of invasive species (see reviews by Buswell et al. 2011, Felker-Quinn et al. 2013, Moran and Alexander 2014), surprisingly few studies have examined the evolutionary consequences of colonization by native species (Foster et al. 2007, Schwarzer et al. 2013, Hargreaves et al. 2014). This knowledge gap is notable because such colonization events are extensive and arguably the most important process in primary and secondary succession, as well as recovery of the Earth's degraded landscapes (Sarrazin and Lecomte 2016).

Invasion biology provides important empirical and theoretical analogs for understanding the evolutionary consequences of range shifts and the colonization of new habitat. Dispersal and colonization are most often considered demographic processes driven by propagule pressure, frequently modeled on simulated landscapes assumed to be homogenous (Travis et al. 2009, Burton et al. 2010). Because few studies have examined native species dynamics, we know much

less about how altered environments drive selection for specific traits and contribute to species persistence in newly created habitats (Jump and Peñuelas 2005, Hargreaves et al. 2014). Evidence of rapid evolution in plants is widespread (Jump and Peñuelas 2005, Lau 2008, Strauss et al. 2008, Buswell et al. 2010, Felker-Quinn et al. 2013), suggesting that successful colonization of novel habitats may result in trait selection and genetically based functional trait shifts relative to the source population.

After the introduction of cane toads in Australia, subsequent colonization of uninvaded habitat led to the evolution of increased dispersal ability along the front of the invasion (Phillips et al. 2006). In the previously described Brazilian system, Zenni et al. (2014a, b) demonstrated rapid evolution of non-native pine trees that escaped silviculture plantations and colonized into native Brazilian ecosystems. These examples point to wide ranging capacity for rapid evolution at the colonizing front, and the rapid evolutionary changes that accompany some invasions can have consequences that alter ecosystems. With the invasive tallow tree (*Sapium sebifera*), new populations of invasive individuals have evolved lower foliar tannins in the invaded range, altering plant herbivore interactions and chemical inputs to soils (Siemann and Rogers 2003). It is changes in traits such as these that can have large subsequent consequences for ecosystem processes.

Plants possess traits that structure belowground communities and influence ecosystem dynamics (Vitousek et al. 1987, Ohtonen et al.1999, Bardgett and Wardle 2010, Cutler et al. 2014). For example, Ohtonen et al. (1999) found that different plant species colonizing a bare soil drastically changed the belowground microbial community. Many studies have shown how invasive species can alter their surroundings by disrupting mycorrhizal relationships (Wolfe et al. 2008), changing soil chemistry (Vitousek et al. 1987, Gómez-Aparicio and Canham 2008), and altering carbon uptake and pool size (see review by Peltzer et al. 2010). Aboveground traits, such as leaf lignin concentrations, can influence belowground nutrient cycles by altering decomposition rates and mineralization of organic matter (Melillo et al. 1982) including within species (Hobbie et al. 2006). Furthermore, intraspecific variation in plant functional traits can alter ecosystem processes (Whitham et al. 2006). For example, leaf litterfall and subsequent decomposition rates were dependent on variation in leaf source within the species *Alnus rubra* (Jackrel and Wootton 2014). Thus, when colonizing trees enter new locations, their traits can

alter belowground ecosystem processes differently than expected from plants growing in established portions of the range (Wardle et al. 2004, Bardgett and Wardle 2010).

Metrosideros polymorpha is considered a foundation species because it is one of the few large trees native to the Hawaiian archipelago and plays an important role in early colonization of lava (Percy et al. 2008, Flaspohler et al. 2010). *Metrosideros polymorpha* is known to have high genetic and phenotypic variation and respond strongly to environmental gradients, such as those that occur across elevation or substrate age (Vitousek 2004, Morrison and Stacy 2014, Stacy et al. 2014). Additionally, genetically based phenotypic variation in *M. polymorpha* has been shown to influence litter decomposition rates and soil nutrient dynamics (Vitousek 2004). The ability to respond to strong environmental gradients and known links between genetically based traits and ecosystem function make *M. polymorpha* an ideal focal plant species for the study of the evolutionary and ecological consequences of colonization (Cordell et al. 1998, Treseder and Vitousek 2001, Martin et al. 2007). To understand the evolutionary and ecological consequences of colonization, we examined how adjacent populations of *Metrosideros polymorpha* varied in functional plant traits and ecosystem processes along a colonization front and strong edaphic and environmental variation in Hawai`i. Using a field and common garden approach, we tested two related hypotheses: 1) functional plant traits within *M. polymorpha* have diverged in newly colonized sites relative to source populations resulting in differential establishment and growth; and 2) variation in traits in *M. polymorpha* in newly colonized sites results in changes to soil processes.

Methods

Study system and sites

The Island of Hawai`i is an ideal location to test the consequences of plant evolution on contemporary timescales. A well-constrained post-volcanic colonization front allows examination of how plant traits differ in newly colonized areas and how these traits influence soil nutrient dynamics (*sensu* Treseder and Vitousek 2001). In 1854-1855 the Mauna Loa volcano erupted, resulting in lava flows that fragmented forests on its eastern face (19.67N, -155.3E). More than 1000 fragments ("kīpuka" in Hawai`ian) were created by the eruption, and range in size from 0.01 to over 100 ha, with large abiotic and biotic environmental differences between the bare substrate of the matrix and the well-developed, 3,000-5,000 yr substrates of

the kīpuka (Raich et al. 1997, Flaspohler et al. 2010, Vaughn et al. 2014, Vannette et al. 2016). These forest fragments persist because primary succession onto new lava is slow – resulting in continuing colonization of the matrix.

The kīpuka-matrix comparison is an ideal field system for studying colonization as an evolutionary process. It is a simple and uniform flora dominated by a single canopy tree species (*Metrosideros polymorpha*), which comprises >85% of the basal area across kīpuka as well as nearly all saplings in the lava flow matrix (Flaspohler et al. 2010, Vaughn et al. 2014). The kīpuka, and the lava flow matrix, are adjacent to one another, thus making these sites ideal for understanding selection of plant functional phenotypes. All kīpuka and matrix sites in this study occur within a narrow geographic area between 1509 and 1637 m above sea level, and so share similar annual ambient temperatures (14.0°C-16.5°C) and precipitation (2400 mm-2900 mm; Western and Juvik 1983, Vaughn et al. 2014), with lava age in this field system not correlated with temperature or rainfall (Tsujii et al. 2016). Lastly, the primary succession of trees onto barren basalt lava flows allows for an examination of the direct effects of plants on organic matter formation and associated soils.

Functional plant traits

To determine if there is variation in functional plant traits and the ecosystem processes they mediate for trees in the colonizing matrix and kīpuka populations, fourteen kīpuka sites surrounded by the adjacent matrix were selected along the Mauna Loa 1855 lava flow. Within each kīpuka site, and also nearby in the surrounding matrix, 10 randomly selected individual trees were sampled and measured in the field. Specific leaf area (SLA), leaf pubescence, and tree height were measured on all trees in the field. Dry leaf mass was determined by oven drying samples at 70° C for 48 h before weighing. Specific leaf area of all kīpuka and matrix trees was determined by calculating the surface area and the mass of three to five leaves per individual (collected from terminal shoots from multiple locations on a tree). Leaf pubescence was estimated in the field with a standardized scale (1-5) by which glabrous leaves were given a score of one and the most pubescent leaves were scored five. Tree heights were estimated to the closest half meter.

Experimental common garden

To determine whether phenotypic differences observed in the field were a plastic response to the environment or resulted from genetic divergence, we established a common garden with rooted cuttings to separate the genetic and environmental components of functional plant traits that vary in response to underlying substrate properties of the kīpuka/lava matrix system (Anderson et al. 2014). Distinct *M. polymorpha* phenotypes were collected from kīpuka and the surrounding lava matrix in June 2012. These cuttings were taken from the same kīpuka sites as the field measurements. However, cuttings from trees on the lava matrix were sampled along a transect spanning the elevation of the kīpuka study site (1509 to 1637 m above sea level), and located between kīpuka (therefore kīpuka and lava cuttings were not paired). Ten 15 cm terminal cuttings from 110 individuals on the lava matrix and 108 individuals in kīpuka were collected in the field and kept moist and cool until planting. Tree cuttings were collected from terminal branch tips and there was no significant difference in cutting diameter between sites or soil substrates (data not shown). Cuttings were scored with pruning shears, and dipped in Hormodin (Indole-3-butyric acid; Hormodin® 2, OHP Inc, Mainland, Pennsylvania, USA). The lower leaves were removed and remaining leaves were cut in half (to reduce water loss), inserted into 1.5 L pots with a standard potting mix (equal parts peat, perlite and vermiculite), and placed under a misting bench (misted every 20 minutes during the day) in a greenhouse facility at the Institute of Pacific Islands Forestry in Hilo, HI, To decrease any potential variation in stored nutrients in the cut branches, successfully rooted cuttings were re-planted in the same potting mix plus an addition of 3 g of 13:13:13 Nitrogen-Phosphorus-Potassium slow release fertilizer pellets one year prior to measuring traits. The greenhouse trees were randomized into four blocks and rotated periodically to avoid any positional environmental effects. In June of 2013, stem diameter and length, specific leaf area (SLA) and leaf pubescence were measured on the new growth. In June of 2014 tree height was measured as the trees were being planted into a common garden in the field at the Institute of Pacific Island Forestry's Laupāhoehoe Science and Education Center (LSEC, Laupāhoehoe, HI, 96743. 19.97094N, -155.24565E) and measured again in 2015. To determine if underlying genetic variation was responsible for differences in tree height, only the individuals who survived to the 2015 measurement were used in the analysis of the 2014 height data. The common garden was designated into 4 random blocks on a site where slope varies from 2 to 15 degrees. Trees were

spaced three meters from each other in a grid and each grid was surrounded by an outer row of edge trees. Within each block a single replicate of each genotype was planted at a random location with ~10 g of NPK 20:20:20 fertilizer in each tree hole.

Plant molecular analyses

To examine genetic structure and gene flow between kīpuka and matrix trees in this system, we used 11 microsatellite markers targeted at repeat regions of the genome (Crawford et al. 2008). Leaves from 168 tree genotypes from the common garden were successfully extracted and genotyped (74 matrix and 94 kīpuka genotypes). Powdered samples of leaf tissue were used to extract genomic DNA (gDNA). Tissues were ground to a fine powder using a ball mill (Spex mixer/mill 8000D, Spex Sample Prep, Metuchen, NJ, 08840). Approximately 0.2 g of leaf powder was used to extract gDNA with the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to manufacturer's instructions, except the first incubation step was conducted overnight (a minimum of 12 hours). Although gDNA yields were low (some less than 5 ng/µL), samples were diluted 1/10 (one part gDNA into nine parts molecular grade water) to minimize the effects of PCR inhibitors for downstream reactions. We generated multi-locus genotypes for each sample using 11 presumably neutral microsatellite markers that were selected from Crawford et al. 2008 (Table S1.1). All PCRs were carried out in 10 µL volumes containing the following reagents (given in final concentrations): 1-5 ng of DNA template, 1x PCR buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 U Platinum *Taq* polymerase (Invitrogen, Carlsbad, CA, USA), and 0.4 µM of each primer, except one marker (MePo514), which was changed to 0.1 µM. A final concentration of 0.5 µg/µL bovine serum albumin (BSA) was added to increase target specificity and PCR yield for four markers (MePo504, MePo507, MePo509 and MePo510). PCRs were thermocycled according to the following conditions: 10 minutes at 95°C to release the Platinum *Taq*® antibody, followed by 40 cycles of 60s at 94°C, 30s at the annealing temperature (T_a) and 30s at 72°C. The T_a, mixing strategy (multiplex vs. singleplex loci and/or pooling scheme), dilution, and forward primer dye for each locus are provided in Supplementary Table 1. Diluted PCR products were electrophoresed on an ABI3130 sequencer with LIZ®-1200 size standard and analyzed using the software GENEMAPPER v4.0 (Applied Biosystems, Foster City, CA, USA). All genotypes were manually checked for accuracy and positive controls were included on all runs. We did not observe amplification in our negative

control reactions (water as template). We also ran independent PCR replicates on 10% of the trees to check for genotyping errors, and no errors were observed.

Soil processes

To determine the variation in soils, we sampled along the constrained colonization sites where macro-environmental conditions are similar (Flaspohler et al. 2010, Tsujii et al. 2016). Soils were collected to a depth up to 20 cm under each of the *M. polymorpha* trees in the kipuka and matrix that were measured for field traits. Soils were collected within 0.25 m from the trunk, placed in plastic bags and stored on ice in a cooler until the end of a day in the field when soils were transferred to a 4° C refrigerator. These soils were almost entirely organic matter, especially in the matrix sites, where soil collection involved scraping a layer of organic material off of basalt bedrock (see supplementary Fig. 1 for detailed explanation and photos of the sites). Samples were shipped on dry ice by 2-day mail to a laboratory at the University of Tennessee where they were stored at 4° C until processed the following day. Processing included sieving soils through a 4 mm mesh and dividing samples for multiple analyses. Soil subsamples were used for fluorometric enzyme assays, soil gravimetric water content, pH, total soil carbon (C) and nitrogen (N), and laboratory incubations to assess C decomposition rates.

Soils were assayed for activities of the following enzymes: β-1,4-glucosidase (EC .2.1.21), α-1,4-glucosidase (EC 3.2.1.20), β-1,4-*N* acetylglucosaminidase (EC 3.1.6.1), acid phosphatase (EC 3.1.3.2), phenol oxidase (EC 1.10.3.2), and peroxidases (EC 1.11.1.7) (Stritar et al. 2010). For these analyses, one gram of each soil was diluted with 125 ml of 50 μM sodium acetate buffer (pH=5) and mixed on a stir plate for 2 minutes, thoroughly suspending the soil in buffer. Enzyme assays were undertaken with eight analytical replicates for β-glucosidase, α-glucosidase, β-1,4-*N* acetylglucosaminidase and phosphatase. Phenol oxidase/peroxidases activities were measured with 16 analytic replicates. The β-glucosidase, α-glucosidase, N acetylglucosaminidase and phosphatase activities were determined by fluorometric response of the 4-methylumbelliferyl (MUB) substrate (excitation at 365 nm, emission at 450 nm, on a BioTek Synergy HT microplate reader, BioTek, Winooski, VT, USA). Phenol oxidase/peroxidase activities were determined by colorimetric analysis of the L-3,4-dihydroxyphenylalanine substrate (DOPA) (Absorbance at 460 nm, SpectraMax Plus^{384} spectrophotometer, Molecular Devices Corp., Sunnyvale, CA, USA). N acetylglucosaminidase

and phosphatase were incubated for 30 min prior to being read, β-glucosidase and α-glucosidase were incubated for 2 hours while phenol oxidase and peroxidase were incubated for 24 hours.

Soil C and N concentrations were determined on finely ground subsamples (mortar and pestle) by dry combustion (Flash EA 1112 CNH analyzer, Thermo Fisher Scientific Inc. 81 Wyman Street, Waltham, MA 02454, USA). Gravimetric water content was assayed by drying soil samples in a Thermo Isotherm soil oven at 105℃ for 48 hours and comparing wet and dry masses. Soil pH was determined using a 0.1M $CaCl_2$ extraction and measured with a Denver Instruments pH probe and reader (Sartorius AG. Weender Landstr. 94-108, 37075 Goettingen, Germany).

Laboratory incubations were conducted over a 30 day period to determine differences in soil C use by microorganisms, and to examine variation in soil C decomposition rates between kīpuka and matrix soils under common conditions (following methods in Schweitzer et al. 2004). Each sieved soil sample was split into two 10 g sub-samples and placed into 125 ml specimen cups and soils were brought up to field capacity (based on field GWC measurements) with the addition of deionized water. All cups were placed into 0.994 L glass jars also containing ~30 ml of deionized water to maintain humidity. The jars were left sealed to incubate in the dark at 22℃. Carbon decomposition rates (g CO2-C / g soil C / day) were measured by comparing carbon dioxide (CO_2) in all jars to a reference at days 1, 2, 4, 8, 14, 22, 27, and 30 by direct injection using an infrared gas analyzer (LI-6400XT, LI-COR Inc, Lincoln NE). Presented carbon decomposition rates represent 30 day cumulative totals.

Statistical analysis

To address whether functional plant phenotypes in the field varied in response to substrate type we took a restricted maximum likelihood approach, predicting height, pubescence, and SLA with substrate location (kīpuka or matrix) as a fixed effect and site as a random effect. Common garden data for these three traits were analyzed separately using a nested analysis of variance with genotype nested within substrate type (kīpuka or matrix). Foliar nitrogen and foliar carbon were also examined as a response to substrate type in the field with a mixed effect model like the one above. Hypothesis testing on all of these mixed effect models was done using a likelihood ratio test between full and null models with an alpha value of 0.05. The R statistical

package was used for all analyses (R Core Team. 2014). The lme4 package (Bates et al. 2015) was used for building statistical models using random effects (varying intercepts).

To understand the potential role of gene flow and natural selection, the common garden quantitative trait differences were compared to microsatellite loci. To determine if the variation in plant phenotype is due to non-random genetic factors, we compared trait variation (Q_{ST}) and genetic variation (F_{ST}), calculated by comparing variation among kīpuka and matrix populations to variation within these populations using the following equation:

$$Qst = \frac{\sigma \mathrm{B}}{\sigma \mathrm{B} + 2\sigma W}$$

where σB is the variance measured among populations and σW is the variance within populations. A significant difference (α=0.05) in average common garden traits between kīpuka plants and matrix plants would allow us to reject the null hypothesis that all variation observed in the field is due to phenotypic plasticity. The quantitative variation was also calculated for each trait to examine the variance accounted for by kīpuka and matrix substrate (Q_{ST}). Secondly, an analysis of molecular variance (AMOVA) was conducted on the 11 microsatellite loci to determine the genetic structuring of the populations. Low variance between populations would suggest that there is little random genetic variance (the null expectation of testing for genetic drift). Comparison of molecular variance at neutral loci (F_{ST}) and quantitative trait variance (Q_{ST}) can discriminate between selection and drift as the driving evolutionary force (Leinonen et al. 2013). For example, a Q_{ST} value greater than an F_{ST} would suggest that trait variation between populations is higher than would be expected by random processes alone. This would suggest that selection on the measured traits is occurring. Paired with a significant difference in the trait value means, we would be able to infer directional selection on plant traits.

To determine if soils vary in response to phenotypic differences among plants, we used a mixed effect model predicting soil C, soil N, soil pH, soil C mineralization with substrate as a fixed effect and site as a random effect. Soil potential enzyme activity data was standardized by soil C before being analyzed between kīpuka and matrix soils like above with a mixed effect model. All of these soil properties were also compared to tree height using an analysis of covariance, where tree height in the field was a continuous variable, substrate type was a categorical variable and the soil chemistry data were continuous responses.

Results

Observations from both the field site and common garden plantings show differences in plant growth traits. Field observations show kīpuka trees were 111.8% taller than matrix trees (Table 1.1, Figure 1.1a, all figures and tables are in the appendix of this chapter). Shoot length, measured in the greenhouse was consistent with field height observations, with shoot lengths in kīpuka-derived plants 53% greater than matrix-derived plants (Figure 1.1b). This pattern was maintained in the common garden where kīpuka trees were 9.2% taller in 2014 (Figure 1.1c) and 9.3% taller than matrix trees in 2015 (Figure 1.1d). These results highlight a consistent pattern of genetic divergence in plant height between the kīpuka and matrix individuals.

Similarly, in the field, we found that other traits differed between trees growing on the two substrate types. Matrix trees were 28.5% more pubescent (Table 1.1) than trees in the kīpuka. Greenhouse pubescence on newly emerged leaves was 17.3% higher in the matrix populations. However, leaf pubescence in the field and greenhouse were weakly correlated, suggesting greater plasticity for this trait. Although *in situ* specific leaf area was observed to be 27.9% greater in kīpuka versus matrix trees, there were no significant differences observed in the common garden after one year of growth, although there appeared to be a trend in this direction. Overall, our quantitative trait analyses support the hypothesis that colonizing trees have unique growth phenotypes, but other traits may be more plastic.

Microsatellite data were used to determine if there was any population genetic structure between the kīpuka and matrix sites. We found no difference in population genetic structure among the kīpuka/matrix pairs (Phi_{PT}=0.001, p=0.46) indicating extensive genetic exchange between the sites (indeed, kīpuka trees likely are the source of matrix populations), and a lack of genetic divergence at neutral loci. Allelic richness was high at many microsatellites (range of 5-49 alleles, mean = 18 alleles per locus) and observed heterozygosity was accordingly high in both kīpuka and matrix populations (H_O=0.67 for both).

As shown above, the kīpuka and matrix populations are genetically indistinguishable at 11 microsatellite loci, suggesting that kīpuka and matrix populations belong to a single interbreeding population. The quantitative trait variance (Q_{ST}) for stem growth in 2013 is 0.323, 0.343 for tree height in 2014, and 0.333 for the tree height in 2015, showing high levels of differentiation between the populations (similar to Alberto et al. 2013) that is much greater than the F_{ST} estimate.

There were five key differences in soil properties associated with trees from kīpuka and matrix sites. Because there was no soil before colonization of the matrix (i.e., the matrix is covered by bare basalt; see Figure S1.1), we make the assumption that effects on soil organic matter in the matrix are due to plant inputs and thus differences in soil properties are due to plants inputs as well as variation in environmental conditions (e.g. temperature, light) that can alter soil microbial activities. Soils collected under trees from kīpuka had higher total soil N (Δ18.2%, Table 1.1, Figure 2a) and total soil C (Δ4.0%, Figure 1.2b) and a 7.2% lower pH (Figure 1.2c). Kīpuka soils were also 10.8% slower at mineralizing C in the lab (Figure 1.2e) that along with the total soil C data indicate higher C storage in kīpuka soils relative to those in the colonizing matrix. Phosphatase activity, when standardized by total soil C, was 47.3% greater in soils collected from under lava matrix plants (Figure 1.2d), likely to utilize the phosphorus (P) bound in the young soils. Gravimetric water content and all other enzyme activities (β-glucosidase, α-glucosidase, N-acetylglucosaminidase, and phenol oxidase) were not significantly different between the matrix and kīpuka soils. These data suggest that overall kīpuka and matrix phenotypes and environmental conditions have differential impacts on belowground processes.

The range of phenotypic variation in growth across all field sites is correlated with variation in soil properties. Despite the coarse estimates of plant growth in the field, changes in soil processes based on tree height independent of substrate were also observed (Table 1.2, Figure 1.2f-j), suggesting that tree phenotypes associated with colonization are altering soils. Soil N, phosphatase activity, and C decomposition rates significantly increased with tree height but did not vary significantly with location, suggesting that plant traits were more important than site conditions in altering these particular soil variables. Soil pH was the only soil trait measured that was significantly affected by the interaction of tree height and location, increasing as trees grew taller in the matrix but not changing in the kīpuka.

Discussion

Our results show that strong edaphic and environmental filters can drive plant genetic divergence and shifts in associated soil processes, despite strong gene flow. Colonization of the new lava matrix substrate resulted in significant genetically based phenotypic changes in functional plant traits, including 9% shorter plants and 17% more pubescence on leaves. A high Q_{ST} value relative to a low F_{ST} for stem growth and height supports the hypothesis that directional selection of colonizing phenotypes is leading to evolution of plant traits on the lava

matrix (Storz 2002, Frei et al. 2014). The reduction in aboveground growth in the matrix is correlated with changes in the belowground ecosystem relative to kīpuka trees, leading to significant decreases in total soil C, N, and acidity, whereas phosphorus availability and carbon decomposition increased. Variation in soil chemistry and microbial function along tree size gradients suggest that tree growth has some level of control over belowground communities, likely via the amount of carbon allocated belowground.

The variation in growth (of new shoots) between kīpuka and matrix trees was consistent in a common garden over two years, suggesting a genetic basis to plant height differences. The alternative explanation of different starting conditions, is minimized because all cuttings were the same length and basal diameter, and thus began growing with the same starting conditions. Although we cannot eliminate other hormonal maternal effects, the fact that the surviving trees showed little variation across time, when measuring traits on new growth, suggests that this effect is minimal or has been stabilized. If the pattern were driven entirely by unequal starting conditions, and not underlying genetic variation, we would expect the difference in heights to continue to decrease.

Soils are a consistent selective filter on plant populations, always interacting with plants, and potentially having dramatic consequences for their evolution. The most obvious examples are serpentine soils, which are hotspots of plant diversity that often contain more trait variation than surrounding locations (Brady et al 2005, Harrison et al. 2006). Soil gradients such as those of mine tailings and serpentine soils are strong selective filters along which evolution of distinct plant phenotypes occur. The difficulty of growing in toxic soils not only leads to novel traits (Brady et al. 2005), but also alters the rates of future evolution of serpentine endemics (Anacker et al. 2010). However, plant evolution across soil gradients is not limited to cases of extreme toxicity, and the evolution of locally adapted phenotypes is commonly due to soil gradients with different available resources (Chapin et al. 1993, Treseder and Vitousek 2001). For example, across a gradient of soil nitrogen in Hawai`i, *M polymorpha* showed distinct genetic separation among sites along with variation in traits associated with nitrogen cycling (Treseder and Vitousek 2001). Furthermore, a recent meta-analysis has shown that species growth response to soil N is better predicted by a phylogenetic approach that incorporates natural selection into models than only incorporating genetic drift (Wooliver et al. 2016). Soils therefore, can have

lasting evolutionary effects on plant traits, not just in extreme examples of toxicity but also along common ecological gradients.

The data shown here show genetically based functional trait variation in an admixed population with complete gene flow, suggesting that sympatric colonization/expansion can also lead to quantitative genetic change. Although invasion is largely an allopatric process (Felker-Quinn et al. 2013), colonization of novel habitats can be sympatric, as seen in expansion fronts (Phillips et al. 2006, Eckert et al. 2008). Unlike allopatric processes, when populations are close geographically, the potential for genetic exchange is high and therefore, natural selection would need to be strong to drive differences among populations. In the examples that exist, it is clear that colonizing individuals are genotypically and phenotypically different than those individuals that exist in the core of the species range (Phillips et al. 2006, Eckert et al. 2008) and that evolution in sympatry may be more common than currently appreciated. For example, a meta-analysis by Eckert et al. (2008) shows that populations at range edges are likely to be more genetically differentiated.

The colonizing individuals in this study demonstrate variation in leaf pubescence and growth during three years in a common garden environment, showing that colonization during primary succession is an evolutionary force even in the face of gene flow. As the Q_{ST}/F_{ST} comparisons in growth traits show, the quantitative trait variation is proportionally greater than neutral genetic variation indicating directional selection on growth (Storz 2002, Leinonen et al. 2013, Frei et al. 2014). With ranges shifting due to natural and anthropogenic causes, there are multiple ways selection can occur. Many studies have shown the evolution of increased growth rate of invaders in invaded habitats (Liao et al. 2013, Matlaga et al. 2012). Increased growth rates at range edges have also been found due to natural range expansion and poleward migration of species (Evans et al. 2013, Kilkenny and Galloway 2013, Schwarzer et al. 2013). Trait evolution on the leading edge of a continuously moving colonization front may maintain a viable colonizing phenotype. The maintenance of distinct ecotypes on a landscape arisen through local adaptation within a species can be attributed to strong edaphic and environmental gradients, and sympatric isolations such as flowering time. For example, dwarf ecotypes of *Eucalyptus globulus* have evolved independently multiple times on rocky cliff outcrops (Foster et al. 2007), but these populations remain much more genetically isolated by distance and phenology, than the rocky colonists *M. polymorpha* studied here.

Ecosystem consequences of trait evolution during colonization

Variation in aboveground plant traits, caused by underlying genetic variation, has been shown to change community and ecosystem processes. Genetic variation within a species can influence associated arthropods (Keith et al. 2010) and soil microbial communities (Schweitzer et al. 2008, Bardget and Wardle 2010) leading to changes in ecosystem function (Hobbie et al. 2006). Furthermore, whether under direct control from plants or indirectly through associated communities, genotypic variation in plant traits has been shown to influence soil respiration (Lojewski et al. 2012) and total soil C and N (Pregitzer et al. 2013) and annual rates of N mineralization (Schweitzer et al. 2012). At broader scales, it has been demonstrated that plant traits influence decomposition rates globally, often being just as important as climate (Cornwell et al. 2008). It is clear that shifts in plant functional traits due to evolution in a novel range can alter ecosystem processes, potentially feeding back on global C and N cycles.

The data reported here suggest that divergence of plant traits, in addition to variation in environmental factors that can alter microbial communities, in matrix trees are significantly changing soil processes in these unique and nutrient poor areas. Variation in soil chemistry such as pH and total N, which are correlated to tree height, as well as differences in light and temperature can alter the microbial communities present and the soil processes they mediate. These different soil communities are acquiring phosphorus at different rates and utilize soil C substrates to different efficiencies, with matrix soils utilizing more recalcitrant C substrates effectively and leading to more C storage in kīpuka beyond the effect of longer storage times. The faster turnover of recalcitrant soil C in the matrix could be due to lower litter quality and the need for specialist microorganisms (Keeler et al. 2009). If this is the case, nutrient limitation in colonizing phenotypes and their associated microbial communities may be decreasing long term C storage within the ecosystem. It is critical to realize that these patterns cannot be separated from the underlying differences in substrate age. However, the soils sampled in this study are very young and comprised primarily of organic inputs. For this reason, we assume that a substantial proportion of the among-site variation in soil chemistry is due to unique plant traits. With potential decreases in tropical C storage in coarse woody debris (Iwashita et al. 2013), increased litter decomposition rates (Bothwell et al. 2014), and belowground process rates (Giardina et al. 2014) as climates warm, understanding how tree genotypes influence C process

rates in tropical fragmented systems will be important to understanding belowground feedbacks to global climate.

Conclusions

Colonization into novel environments occurs constantly in both natural and anthropogenically driven contexts, with colonizing species evolving due to biotic and abiotic filters encountered in the new habitats. Our results from the field, the greenhouse and common garden measurements show that divergence in growth occurs, despite strong gene flow, and divergence can lead to variation in growth and other functional traits. Moreover, these data show that phenotypic differences, in combination with environmental differences across the sites, may alter soil properties and ecosystem processes. These data support previous work showing that plant colonization may lead to niche construction, creating distinct soil conditions that influence soil C and nutrient dynamics. Foundation species can rapidly evolve and the ecosystem consequences of these colonizing phenotypes are critical to understanding the full effects of plant species migration under both natural and anthropogenic circumstances.

The Hawai`ian Islands provide globally unique model study systems for testing ecological theory (Vitousek 2004), and understanding evolutionary change, including radial divergence (Freed et al. 1987), rapid evolution (Carson and Johnson 1975), and phenotypic plasticity (Cordell et al. 1998). Re-colonization of lava by *M. polymorpha* post-eruption provides another model study system for understanding the long term evolutionary consequences of colonization, especially as this tree species is both an early colonizer and a long-lived canopy dominant. This implies that the strong abiotic changes that are encountered during colonization can act as a selective agent. This study supports the idea that evolutionary processes in land plants can be rapid along soil gradients (Chapin et al. 1993. Treseder and Vitousek 2001, Brady et al. 2005, Buswell et al. 2011). The drastic change in the potential fitness between the kīpuka and lava matrix has likely led to traits being selected for during colonization of the lava matrix. Colonization is a subset of species movement, similar to invasion, in both how plants evolve in novel locations, and how evolved phenotypes alter ecosystems. Drawing on both of these ideas in a broad framework of species movement is necessary for understanding ecosystem consequences in a changing world.

Acknowledgements

We thank Chris Wong, Ashley Shaw, Alysia Kaha, Riley DeMattos, James Akau, Kainana Francisco, Michael Van Nuland, Greg Newman, Caroline Daws and Peter Meidl for assistance in the field and laboratory. We are grateful to the U.S. Forest Service: Institute for Pacific Island Forestry for greenhouse space and the land for planting the common garden in Hawai`i. This work was funded by the University of Tennessee, the USDA Forest Service Pacific Southwest Research Station, and the Americorp Program managed by KUPU, Hawai`i, the Cowden Endowment at Northern Arizona University, National Science Foundation grant DEB–1020412 (T. Fukami, C. P. Giardina, G. P. Asner), and the Hooper Undergraduate Research Award.

References

Alberto, F.J., S.N. Aitken, R. Alía, S.C. González-Martínez, H. Hännin, A. Kremer, F. Lefèvre, T. Lenormand, S. Yeaman, R. Whetten, and O. Savolainen. 2013. Potential for evolutionary responses to climate change – evidence from tree populations. Global Change Biology 19:1645-1661.

Anacker B.L, J.B. Whittall, E.E. Goldberg, and S.P. Harrison. 2010. Origins and consequences of serpentine endemism in the California flora. Evolution 65:365-376.

Anderson J.T., M.R. Wagner, C.A. Rushworth, K.V.S.K. Prasad, and T. Mitchell-Olds. 2014. The evolution of quantitative traits in complex environments. Heredity 112: 4-12.

Bailey, J.K., M.A. Genung, I. Ware, C. Gorman, M.E. Van Nuland, H. Long, and J.A. Schweitzer. 2014. Indirect genetic effects: an evolutionary mechanisim linking feedbacks, genotypic diversity and coadaptation in a climate change context. Functional Ecology 28:87-95.

Bardgett, R.D., and D.A. Wardle.2010. Aboveground-belowground linkages: biotic interactions, ecosystem processes, and global change. in Harvery P.H., R.M. May, C.J. Godfray, J.A. Dunne, editors. Oxford Series in Ecology and Evolution. Oxford University Press, New York, USA.

Bates D., M. Maechler, B. Bolker, S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1-48.

Baruch, Z., and G. Goldstein. 1999. Leaf construction cost, nutrient concentration, and net CO2 assimilation of native and invasive species in Hawaii. Oecologia 121:183-192.

Bothwell, L.D., P.C. Selmants, C.P. Giardina, and C.M. Litton. 2014. Leaf litter decomposition rates increase with rising mean annual temperature in Hawaiian tropical montane wet forests. PeerJ 2:685.

Brady, K.U., A.R Kruckeberg, and H.D.Bradshaw Jr. 2005. Evolutionary ecology of plant adaptation to serpentine soils. Annual Review of Ecology, Evolution, and Systematics 36:243-266.

Burton, O.J., B.L. Phillips, and J.M.J. Travis. 2010. Trade-offs and the evolution of life-histories during range expansion. Ecology Letters 13:1210-1220.

Buswell, J. M., A. T. Moles, and S. Hartley. 2011. Is rapid evolution common in introduced plant species? Journal of Ecology 99:214-224.

Carroll, S.P., J.E. Loye, H. Dingle, M. Mathieson, T.R. Famula, and M.P. Zalucki. 2005. And the beak shall inherit–evolution in response to invasion. Ecology Letters 8:944-951.

Carson, H.L. and W.E. Johnson. 1975. Genetic variation in hawaiian Drosophila I. Chromosome and allozyme polymorphism in D.setosimentum and D. ochrobasis from the Island of Hawaii. Evolution 29:11-23.

Chapin III, S.F., K. Autumn, and F. Pugnaire. 1993. Evolution of suites of traits in response to environmental stress. The American Naturalist 142:S78-S92.

Clausen. J, D.D. Keck, and W.H. Hiesey. 1947. Heredity of geographically and ecologically isolated races. The American Naturalist 81:114-133.

Cordell, S., G. Goldstein, D. Mueller-Dombois, D. Webb, and P. M. Vitousek. 1998. Physiological and morphological variation in Metrosideros polymorpha, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. Oecologia 113:188-196.

Cornwell, W.K., J.H. Cornelissen, K. Amatangelo, E. Dorrepaal, V.T. Eviner, O. Godoy, S.E. Hobbie, B. Hoorens, H. Kurokawa, N. Pérez-Harguindeguy, H.M. Quested, L.S. Santiago, D.A. Wardle, I.J. Wright, R. Aerts, S.D. Allison, P. Van Bodegom, V. Brovkin, A. Chatain, T.V. Callaghan, S. Díaz, E. Garnier, D.E. Gurvich, E. Kazakou, J.A. Klein, J. Read, P.B. Reich, N.A. Soudzilovskaia, M.V. Vaieretti, and M. Westoby. 2008. Plant species traits are

the predominant control on litter decomposition rates within biomes worldwide. Ecology letters 11:1065-1071.

Crawford, N.G., C. Hagen, H.F. Sahli, E.A. Stacy, and T.C. Glenn. 2008. Fifteen polymorphic microsatellite DNA loci from Hawaii's Metrosideros polymorpha (Myrtaceae: Myrtales), a model species for ecology and evolution. Molecular Ecology Resources 8:308–310

Cutler, N. A., D. L. Chaput, and C. J. van der Gast. 2014. Long-term changes in soil microbial communities during primary succession. Soil Biology and Biochemistry 69:359-370.

Eckert, C.G., K.E. Samis, and S.C. Lougheed. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17:1170-1188.

Evans, G. A., F. F. Kilkenny, and L. F. Galloway. 2013. Evolution of competitive ability within Lonicera japonica's invaded range. International Journal of Plant Sciences 174:740-748.

Felker-Quinn, E ., J. A. Schweitzer, and J. K. Bailey. 2013. Meta-analysis reveals evolution in invasive plant species but little support for Evolution of Increased Competitive Ability (EICA). Ecology and Evolution 3:739-751.

Flaspohler, D. J., C. P. Giardina, G. P. Asner, P. Hart, J. Price, C. K. a. Lyons, and X. Castaneda. 2010. Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. Biological Conservation 143:280-288.

Foster, S.A., G.E. McKinnon, D.A. Steane, B.M. Potts, and R.E. Vaillancourt. 2007. Parallel evolution of dwarf ecotypes in the forest tree Eucalyptus globulus. New Phytologist 175:370-380.

Freed, L.A., S. Conant, and R.C. Fleischer. 1987. Evolutionary ecology and radiation of Hawaiian passerine birds. Trends in Ecology and Evolution 2:196-203.

Frei, E.R., T. Hahn, J. Ghazoul, and A.R. Pluess. 2014. Divergent selection in low and high elevation populations of a perennial herb in the Swiss Alps. Alpine Botany 124:131-142.

Giardina, C.P., C.M. Litton, S.E. Crow, and G.P. Asner. 2014. Warming-related increases in soil CO2 efflux are explained by increased below-ground carbon flux. Nature Climate Change 4:822-827

Gómez-Aparicio, L., and C. D. Canham. 2008. Neighborhood models of the effects of invasive tree species on ecosystem processes. Ecological Monographs 78:69-86.

Hargreaves, A.L., K.E. Samis, and C.G. Eckert. 2014. Are species' range limits simply

niche limits writ large? A review of transplant experiments beyond the range. The American Naturalist 183:157-173.

Harrison, S., H.D. Safford, J.B. Grace, J.H. Viers, and K.F. Davies. 2006. Regional and local species richness in an insular environment: serpentine plants in California. Ecological Monographs 76:41-56.

Hobbie, S.E., P. B. Reich, J. Oleksyn, M. Ogdahl1, R. Zytkowiak, C. Hale, and P. Karolewski. 2006. Tree species effects on decomposition and forest floor dynamics in a common garden. Ecology 87:2288–2297.

Hobbs, R. J., S. Arico, J. Aronson, J. S. Baron, P. Bridgewater, V. A. Cramer, P. R. Epstein, J. J. Ewel, C. A. Klink, and A. E. Lugo. 2006. Novel ecosystems: theoretical and management aspects of the new ecological world order. Global ecology and biogeography 15:1-7.

Holt, R. D. 2003. On the evolutionary ecology of species' ranges. Evolutionary Ecology Research 5:159-178.

Iwashita, D.K., C.M. Litton, and C. P. Giardina. 2013. Coarse woody debris carbon storage across a mean annual temperature gradient in tropical montane wet forest. Forest Ecology and Management 291:336-343.

Jump, A.S. and J. Peñuelas. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters 8:1010-1020.

Keeler, B. L., S.E. Hobbie, and Kellog L.E. 2009. Effects of long-term nitrogen addition on microbial enzyme activity in eight forested and grassland sites: Implications for litter and soil organic matter decomposition. Ecosystems 12:1-15.

Keith, A. R., J.K. Bailey, and T.G. Whitham. 2010. A genetic basis to community repeatability and stability. Ecology 91:3398-3406.

Kilkenny, F.F., and L.F. Galloway. 2013. Addaptive divergence at the margin of an invaded range. Evolution 67:722-731.

Kinnison, M.T., N.G. Hairston, and A.P. Hendry. 2015. Cryptic eco-evolutionary dynamics. Annals of the New York Acadamy of Sciences 1360:120-144

Lau, J. A. 2008. Beyond the ecological: biological invasions alter natural selection on a native plant species. Ecology 89:1023-1031.

Liao, Z. Y., R. Zhang, G. F. Barclay, and Y. L. Feng. 2013. Differences in competitive ability between plants from nonnative and native populations of a tropical invader relates to adaptive responses in abiotic and biotic environments. PloS One, 8, e71767.

Leinonen, T., R.J. McCairns, R.B. O'Hara, and J. Merilä. 2013. Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics 14:179-190.

Lindström, T.,G.P. Brown, S.A. Sisson, B.L. Phillips, and R. Shine. 2013. Rapid shifts in dispersal behavior on an expanding range edge. Proceedings of the National Academy of Sciences 110:13452-13456.

Lojewski, N. R.,D.G. Fischer, J.K. Bailey, J.A. Schweitzer,T.G. Whitham, and S.C. Hart, 2012. Genetic components to belowground carbon fluxes in a riparian forest ecosystem: a common garden approach. New Phytologist 195:631-639.

Martin, R. E., G. P. Asner, and L. Sack. 2007. Genetic variation in leaf pigment, optical and photosynthetic function among diverse phenotypes of Metrosideros polymorpha grown in a common garden. Oecologia 151:387-400.

Matlaga, D. P., L. D. Quinn, A. S. Davis, and J.R. Stewart, 2012. Light response of native and introduced Miscanthus sinensis seedlings. Invasive Plant Science and Management 5: 363-374.

Melillo, J.M., J. D. Aber, and J. F. Muratore. 1982. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. Ecology 63:621-626.

Moran, E.V., J.M. Alexander. 2014. Evolutionary responses to global change: lessons from invasive species. Ecology letters 17:637-649.

Morrison, K. R., and E.A. Stacy. 2014. Intraspecific divergence and evolution of a life-history trade-off along a successional gradient in Hawaii's Metrosideros polymorpha. Journal of Evolutionary Biology 27:1192-1204.

Ohtonen, R., H. Fritze, T. Pennanen, A. Jumpponen, and J. Trappe. 1999. Ecosystem properties and microbial community changes in primary succession on a glacier forefront. Oecologia 119:239-246.

Parmesan, C. 2006. Ecological and evolutionary responses to recent climate change. Annual Review of Ecology, Evolution, and Systematics 37:637-669.

Peltzer, D. A., R. B. Allen, G. M. Lovett, D. Whitehead, and D. A. Wardle. 2010. Effects of biological invasions on forest carbon sequestration. Global Change Biology 16:732-746.

Percy, D.M., A. M. Garver, W. L. Wagner, H. F. James, C. W. Cunningham, S. E. Miller, R. C. Fleischer. 2008. Progressive island colonization and ancient origin of Hawaiian Metrosideros (Myrtaceae). Procedings of the Royal Society B 275:1479-1490.

Phillips, B.L., G. P. Brown, J. K. Webb and R. Shine. 2006. Invasion and the evolution of speed in toads. Nature 439:803

Pregitzer, C.C., J.K. Bailey, and J.A. Schweitzer. 2013. Genetic by environment interactions affect plant–soil linkages. Ecology and Evolution 3:2322-2333.

Prentis, P.J., J.R.U. Wilson, E. E. Dormontt, D. M. Richardson, and A. J. Lowe. 2008. Adaptive evolution in invasive species. Trends in Plant Science 13:288-294

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raich, J.W., A.E. Russell, P.M. Vitousek. 1997. Primary productivity and ecosystem development along an elevational gradient on Mauna Loa, Hawai'i. Ecology 78:707-721

Read, Q.D., L.C. Moorhead, N.G. Swenson, J.K. Bailey, and N.J. Sanders. 2014. Convergent effects of elevation on functional leaf traits within and among species. Functional Ecology 28:37-45.

Sarrazin, F., and J. Lecomte. 2016. Evolution in the Anthropocene. Science 351:922-923.

Schwarzer, C., T. Heinken, V. Luthardt, and J. Joshi. 2013. Latitudinal shifts in species interactions interfere with resistance of southern but not of northern bog-plant communities to experiemental climate change. Journal of Ecology 101:1484-1497.

Schweitzer, J.A., J.K. Bailey, B.J. Rehill, G.D. Martinsen, S.C. Hart, R.L. Lindroth, P. Keim, and T.G. Whitham. 2004. Genetically based trait in a dominant tree affects ecosystem processes. Ecology letters 7:127-134.

Schweitzer, J. A., J. K. Bailey, S. C. Hart, and T. G. Whitham. 2005. Nonadditive effects of mixing cottonwood genotypes on litter decomposition and nutrient dynamics. Ecology 86:2834-2840.

Schweitzer, J. A., J. K. Bailey, D. G. Fischer, C. J. LeRoy, E. V. Lonsdorf, T. G. Whitham, and S. C. Hart. 2008. Plant-soil-microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms. Ecology 89:773-781.

Schweitzer J.A., M.D. Madritch, E. Felker-Quinn, J.K.Bailey. 2012. From genes to ecosystems: plant genetics link above- and belowground processes. in Soil Ecology and Ecosystem Services, eds Wall DH, et al. (Oxford University Press), pp.82–97.

Schweitzer J.A., I. Juric, T.F.J. Voorde, K. Clay, W.H. Putten, and J.K. Bailey. 2014. Are there evolutionary consequences of plant-soil feedbacks along soil gradients? Functional Ecology 28:55-64.

Siemann, E., and W.E. Rogers. 2003. Herbivory, disease, recruitment limitation, and success of alien and native tree species. Ecology 84:1489-1505.

Stacy, E.A., J.B. Johansen, T. Sakishima, D.K. Price, and Y. Pillon. 2014. Incipient radiation within the dominant Hawaiian tree Metrosideros polymorpha. Heredity 113:334-342.

Storz, J.F. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. Molecular Ecology 11:2537-2551

Strauss, S. Y., J. A. Lau, T. W. Schoener, and P. Tiffin. 2008. Evolution in ecological field experiments: implications for effect size. Ecology Letters 11:199-207.

Stritar, M.L., J.A. Schweitzer, S.C. Hart, and J.K. Bailey. 2010. Introduced ungulate herbivore alters soil processes after fire. Biological Invasions 12:313-324.

Travis, J.M.J., K. Mustin, T.G. Benton, and C. Dytham. 2009. Accelerating invasion rates result from the evolution of density-dependent dispersal. Journal of Theoretical Biology 259:151-158

Treseder, K.K., and P.M. Vitousek. 2001. Potential ecosystem-level effects of genetic variation among populations of Metrosideros polymorph from a soil fertility gradient in Hawaii. Oecologia 126:266-275

Tsujii, Y., Y. Onoda, A. Izuno, Y Isagi, and K. Kitayama. 2016. A quantitative analysis of phenotypic variations of Metrosideros polymorpha within and across populations along environmental gradients on Mauna Loa, Hawaii. Oecologia 180:1049-1059

Vannette, R. L., D.R. Leopold, T. Fukami. 2016. Forest area and connectivity influence root-associated fungal communities in a fragmented landscape. Ecology 97:2374-2383.

Vaughn, N.R., G.P. Asner, and C.P. Giardina. 2014. Centennial impacts of fragmentation on the canopy structure of tropical montane forest. Ecological Applications 24:1638-1650.

Vitousek, P. M., L. R. Walker, L. D. Whiteaker, D. Mueller-Dombois, and P. A. Matson. 1987. Biological Invasion by Myrica faya alters ecosystem development in Hawaii. Science 238:802-804.

Vitousek, P.M. 2004. Nutrient cycling and limitation: Hawai'i as a model system. Princeton University Press.

Wardle, D.A., R.D. Bardgett, J.N. Klironomos, H. Setälä, W.H. van der Putten, and D.H. Wall. 2004. Ecological linkages between aboveground and belowground biota. Science 304:1629-1633.

Western, L., and J.O. Juvik. 1983. Roadside plant communities on Mauna Loa, Hawaii. Journal of Biogeography 1:307-316.

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. LeRoy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, and B. M. Potts. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics 7:510-523.

Wolfe, B. E., V. L. Rodgers, K. A. Stinson, and A. Pringle. 2008. The invasive plant Alliaria petiolata (garlic mustard) inhibits ectomycorrhizal fungi in its introduced range. Journal of Ecology 96:777-783.

Wooliver R., A.A. Phennigwerth, J.K. Bailey, and J.A. Schweitzer. 2016. Plant functional traits constraints guide macroevolutionary trade-offs in competitive and conservitive growth responses to nitrogen. Functional Ecology 30:1099-1108

Zenni, R.D., J. Lamy, L. J. Lamarque, and A. J. 2014a. Porté Adaptive evolution and phenotypic plasticity during naturalization and spread of invasive species: implications for tree invasion biology. Biological Invasions 16:635-644

Zenni, R.D. J.K. Bailey, and D. Simberloff. 2014b. Rapid evolution and range expansion of an invasive plant are driven by provenance–environment interactions. Ecology Letters 17:727-735

Appendix

Table 1.1: Variation in plant traits in the field and common garden. Analysis of variance for field observations (Field 2012) and related common garden traits (CG 2013-2015). P-values (p) are shown along with the percent difference between the traits from trees from the kīpuka and surrounding lava matrix. Negative percent differences represent situations where the trait or response is greater in the kīpuka and positive percent differences represent higher matrix values. Growth in the field was estimated with height, while shoot length was used to estimate growth in the greenhouse. Growth measurements were taken in a common environment in Hawai`i for three consecutive years; 2013 in the greenhouse and 2014 and 2015 in the outplanted common garden. Bolded p-values represent significant effects at α=0.05.

Response	p-value	% change in matrix
Plant Functional Traits		
Pubescence	**<0.0001**	25.8%
Height	**<0.0001**	-111.8%
Specific Leaf Area	**<0.0001**	-27.9%
Chemistry		
Foliar Nitrogen	.99	0
Foliar Carbon	**<0.0001**	1.6%
Soil total Nitrogen	**<0.0001**	-18.2%
Soil total Carbon	**0.0071**	-4.0%
Soil pH	**<0.0001**	7.2%
Soil CO_2 incubation	**0.0495**	10.8%
β-glucosidase	0.0941	33.8%
α-glucosidase	0.4565	33.6%
N-acetylglucosaminidase	.8838	0.67%
Phosphatase	**0.0192**	47.3%
Phenol oxidase	0.0603	159.4%
Plant Functional Traits		
Pubescence 2013	**<0.0001**	17.3%
Specific Leaf Area 2013	0.0917	-2.5%
Stem diameter 2013	0.1383	-3.2%
Shoot length 2013	**<0.0001**	-46.9%
Height 2014	**0.0178**	-9.2%
Height 2015	**0.0424**	-9.3%

Table 1.2: Variation in tree height and substrate influence mortality in the common garden and soil properties in the field. Analysis of covariance for field soil response variables to tree height (Height) and the soil age (Substrate). F ratios for each parameter are listed followed by the p value in parentheses. Bolded F ratios represent significant (α=0.05) effects and italicized F ratios represent p <0.1.

Response	Height	Substrate	Height X Substrate
Soil chemical data			
Soil total Nitrogen	**25.14**(<.0001)	1.1902 (0.28)	0.3473 (0.55)
Soil total Carbon	0.9459 (0.33)	1.6304 (0.20)	0.1137 (0.74)
Soil pH	2.3870 (0.16)	2.5363 (0.11)	**5.1746** (0.03)
Soil enzyme activities			
Phosphatase	**6.3877** (0.01)	1.3947 (0.24)	0.0769 (0.78)
β-glucosidase	0.5874 (0.44)	2.3661 (0.13)	0.0229 (0.88)
α-glucosidase	1.0004 (0.32)	0.0503 (0.82)	0.0116 (0.91)
NAG	1.6434 (0.20)	0.0172 (0.90)	0.1146 (0.74)
Phenol oxidase	*3.1962* (0.08)	1.5699 (0.21)	0.3323 (0.57)
Soil CO_2 incubation			
30 d decomposition	**4.3479** (0.04)	0.5959 (0.44)	0.1534 (0.70)

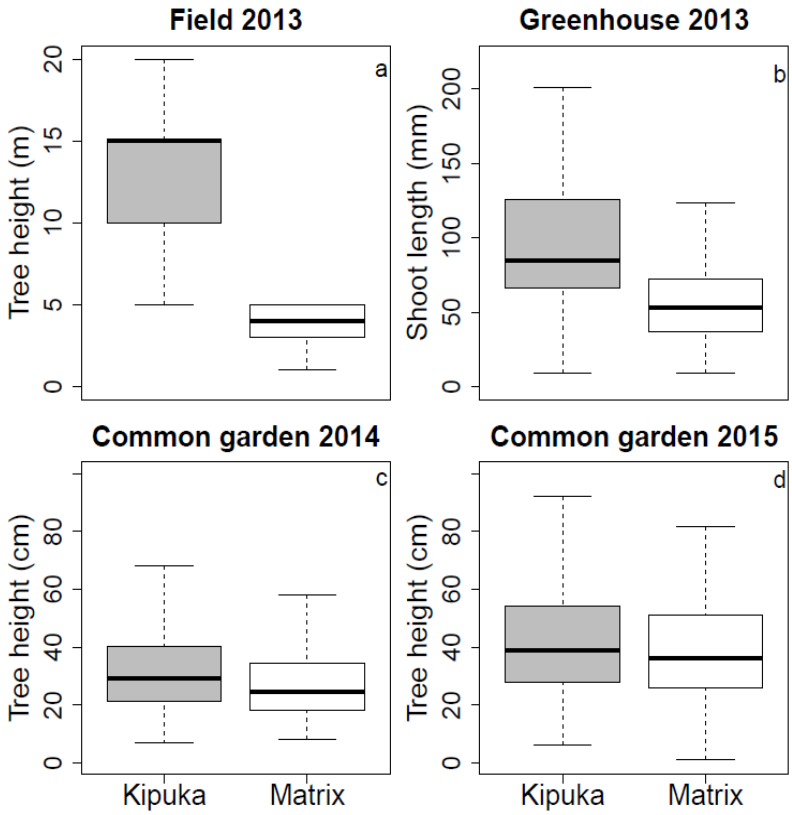


Figure 1.1: Colonization results in variation in plant growth traits. In both the field (a) and a common garden experiment across years (b-d), trees that colonized the lava matrix have significantly reduced growth (i.e., shorter total heights in field/common garden and shoot length in the greenhouse). Boxes represent the distribution of the first to third quartiles of the data while tails represent 95% confidence intervals. Each panel is showing a significant difference between means at an alpha value $\alpha = 0.05$.

Figure 1.2: Variation in soil properties between kīpuka and matrix substrates. Soils in the older substrate had higher concentrations of (a) carbon (C), (b) nitrogen (N), and (c) were more acidic. Soils in the younger substrate had (d) higher phosphatase activities and (e) higher C decomposition. Increases in tree height and shoot length, significantly increases (f) soil N found beneath these trees but not (g) carbon. Soil pH (h, black solid line) significantly increases with height in the matrix, without (h, grey dashed line) changing in the kīpuka. Increases in the (i) extracellular enzyme activity of phosphatase were found to increase in soils of shorter trees, and soils under shorter trees also have increased (j) C decomposition (i.e., reduced soil C storage), when measured during a 30 day lab incubation.

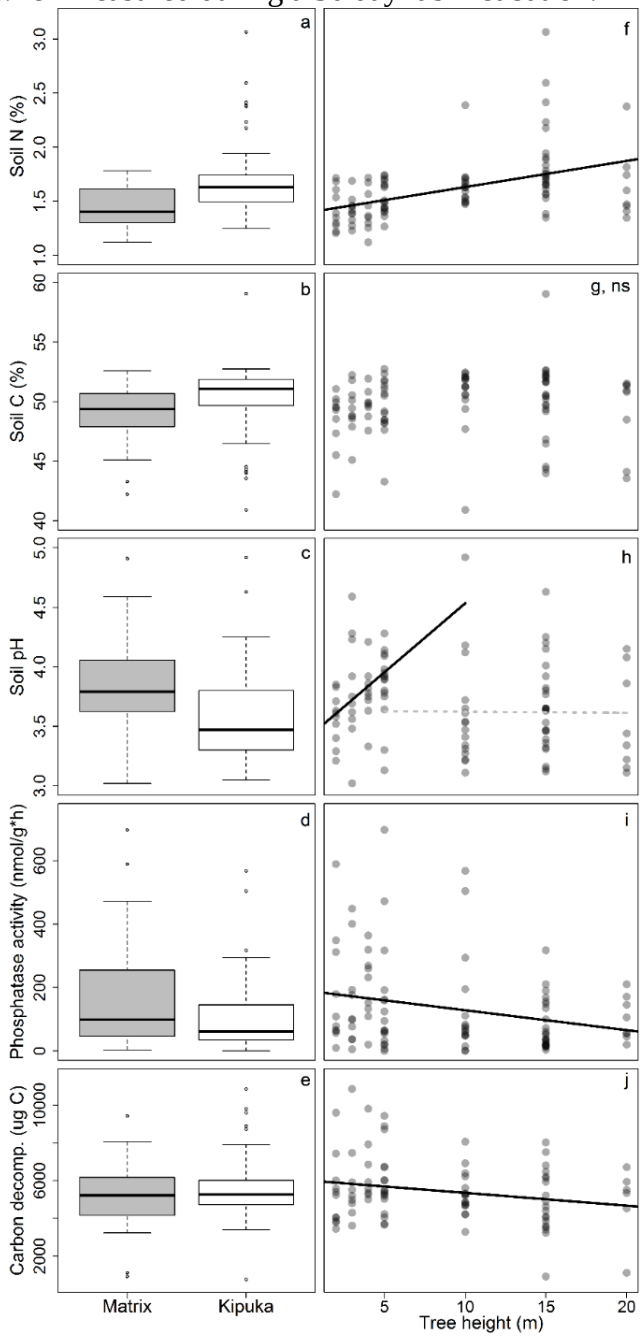


Table S1.1: Microsatellite markers used to estimate population structure in *Metrosideros polymorpha*.
T_a is the annealing temperature in PCR; *A* is the total number of observed alleles. Loci were amplified in duplexes or as singletons; singletons were pooled together or added to Mix 1 before being run on an AB3130.

Marker	T_a	**Mix**	**Post-PCR dilution**	**Dye**	***A***	**Allele sizes (bp)**	**Primer sequence (5'-3')** (Crawford et al. 2008)
MePo501	55	Pool1	1/80	NED	14	102, 115, 116, 117, 118, 125, 126, 128, 130, 132, 134, 138, 140, 142	F-TCTTTCGCCGGATTACTT R-GAGTGCCTTATTCATGCTATGT
MePo503	60	Pool into Mix 1	1/60	6FAM	44	207, 215, 219, 223, 231, 235, 239, 243, 247, 251, 255, 259, 263, 267, 271, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319, 323, 327, 328, 331, 335, 339, 340, 345, 349, 353, 357, 361, 365, 369, 373, 377, 381	F-CTCACATCGCTTGTCTA R-CCAAATTAAGAACGATACAT
MePo504	55	Pool1	1/25	6FAM	10	128, 132, 136, 140, 144, 148, 152, 156, 160, 164	F-AGGGGGATTGTTGGTTAT R-AGAGGAATGGGAAATAAC
MePo506	60	Pool into Mix 1	1/60	VIC	49	218, 222, 226, 232, 238, 244, 250, 256, 260, 262, 266, 272, 278, 284, 290, 296, 300, 306, 312, 318, 324, 330, 336, 342, 346, 352, 358, 364, 370, 376, 380, 386, 392, 398, 404, 410, 416, 420, 426, 432, 438, 442, 448, 454, 466, 470, 486, 490, 502	F-ATCCCCTCACGATTATAG R-ATGGGCTGACGAATATA
MePo507	55	Pool1	1/80	NED	19	214, 216, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 250, 254, 256	F-TTTTCTTGGCAYGACTCT R-CGTCTACAAATCCAAGTAAA
MePo508	50	Pool1	1/40	VIC	5	182, 186, 190, 194, 198	F-ACTTATTGAGTGTGACCTATAC R-TAGAGCGAAGCAAATATAT
MePo509	48	Pool1	1/4	6FAM	9	173, 177, 181, 185, 189, 193, 201, 205, 207	F-GTTCGCGGTTGATTACTA R-GAATCACAAGCCATTAG

Table S1.1: Continued

Marker	Ta	Mix	Post-PCR dilution	Dye	A	Allele sizes (bp)	Primer sequence (5'-3') (Crawford et al. 2008)
MePo510	48	Pool into Mix 1	1/6	PET	16	214, 220, 224, 226, 228, 230, 232, 234, 236, 238, 240, 244, 248, 252, 256, 260	F-CGAAACGGATCAAGATAGT R-CTTTTTGGGGACAATTACTA
MePo512	55	Pool into Mix 1	1/30	NED	43	145, 157, 161, 165, 169, 173, 177, 181, 185, 189, 193, 197, 201, 203, 205, 209, 211, 213, 215, 217, 221, 223, 225, 227, 229, 233, 237, 239, 241, 245, 249, 255, 257, 261, 265, 273, 277, 279 ,281, 287, 295, 303, 309	F-ACGAATTGCTTTATTGATATAC R-TAAATTGGGCCACATAGA
MePo513	60	1	1/15	VIC	8	186, 190, 194, 198, 200, 202, 204, 206	F-CAGAACTGGTATCCTGATATA R-GTTTAATGAGTTGGGATTTAGAG
MePo514	60	1	1/15	PET	16	162, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194	F-GCTTTCCTGCAACAGAGTAG R-GTGCATCAACCGCTACAT

Loci listed as "pool" in the Mix column are run singly in separate PCRs, and then diluted and either pooled together (Pool 1) or added to Mix 1 for loading onto an ABI3130

Figure S1.1: Explanation and photos of field sites. In 1854-1855 the Mauna Loa volcano erupted, and covered the landscape with basalt, leaving small patches of intact forest. The trees colonizing the lava matrix grow directly into the basalt. The soils collected in the matrix are very young (under 180 years) and appear to be primarily comprised of decaying organic matter from the tree above (**a**). The lack of measurable soil horizons is further shown in the next panel (**b**) where a tree has recently fallen, exposing the basalt substrate directly beneath the roots.

CHAPTER II

POPULATIONS OF *POPULUS ANGUSTIFOLIA* HAVE EVOLVED DISTINCT METABOLIC PROFILES THAT INFLUENCE THEIR SURROUNDING SOIL

A version of this chapter is in review at *Plant and Soil* and was authored by: Liam O Mueller, Samuel R. Borstein, Eric D. Tague, Stephen P. Dearth, Hector F. Castro, Shawn R. Campagna, Joseph K. Bailey, and Jennifer A. Schweitzer

Liam O Mueller, Samuel R. Borstein, Eric D. Tague, Stephen P. Dearth, Hector F. Castro, Shawn R. Campagna, Joseph K. Bailey, and Jennifer A. Schweitzer (2019). Populations of *Populus angustifolia* have evolved distinct metabolic profiles that influence their surrounding soil. Plant and Soil (*In review*)

LOM, HFC, SRC, JKB, and JAS formulated the initial questions. EDT, SPD, HFC, and SRC performed the chemical assays. LOM, SRB, EDT, HFC, and SRC performed the chemical analysis. LOM and SRB performed the statistical analyses. Writing was shared by all authors. All figures were moved to the appendix of this chapter.

Abstract

Plant-microbial-soil interactions are key to understanding plant community succession, invasion success, patterns of biodiversity and aspects of ecosystem function. Yet root and rhizosphere chemistry is highly complex, and little is known about natural variation across environmental gradients. Here, we used metabolomics to assess bulk small molecule profiles addressing the hypothesis that genetic variation across a species range would result in varying metabolic profiles. Using UPLC-HRMS we assessed the small molecule profile of root tissue and surrounding rhizosphere soil from 5 year old plant clones collected from six populations of *Populus angustifolia* across the western U.S., grown in a common environment. Population-level variation was found in root metabolomes and soil organic composition across the populations. Redundancy analysis of over twelve thousand metabolites suggest that plant population origin can account for up to 30% of the variation in roots and 27% of the variation in rhizosphere soil chemistry. Co-inertia analysis indicates that variation in root metabolite profiles explains 15% of the variation in paired soil samples. Distinct populations have evolved different root tissue metabolomes. The difference in root metabolites across populations altered the rhizosphere soil composition in unique ways, signaling the plant conditioning of soil varies by plant population.

Introduction

Plant traits alter soil properties and structure belowground communities, both of which can have ecological and evolutionary consequences for plant populations, community dynamics, and ecosystem function (Vitousek et al. 1987, Whitham et al. 2006, Bardgett and Wardle 2010). Plant-soil linkages and feedbacks are an important process that can drive plant ranges and distribution, (Bezemer et al. 2006, van der Putten et al. 2016, Van Nuland et al. 2017), improve invasion success (Van der Putten 2003, Wolfe et al. 2008), and hasten succession (Kardol et al. 2006). For example, aboveground traits, such as leaf lignin concentration, can influence belowground nutrient cycles by altering decomposition rates and mineralization of organic matter (Melillo et al. 1982)(Hobbie et al. 2006). Furthermore, whether under direct control from plants or indirectly through associated communities, genotypic variation in plant traits has been shown to influence a broad range of ecosystem processes a few of which include: soil respiration (Lojewski et al. 2012), total soil carbon and nitrogen (Pregitzer et al. 2013), ammonium availability (Madritch and Hunter 2005), annual rates of N mineralization (Schweitzer et al. 2011), extracellular enzyme activity (Madritch et al. 2009), and litter decomposition (Crutsinger et al. 2009). Plant phenotypes belowground can alter soils (Bardgett and Wardle 2010), exuding compounds that can alter microbial communities (Badri et al. 2009, Rasmann and Turlings 2016), direct biotic interactions such as root herbivory, and mediate communication between plants and microbes (Hu et al. 2018). For example, the nitrogen cycle, is more directly controlled by plant root exudates than previously thought (Coskun et al. 2017) with root compounds acting as biological nitrogen uptake inhibitors, denitrification inhibitors, as well as mycorrhizal signals. These secondary plant metabolites play a key role in plant-soil interactions, but it is only recently that comprehensive exploration of the complex chemical interactions between the plant and rhizosphere has been attempted (van Dam and Bouwmeester 2016, Mhlongo et al. 2018).

Metabolomics is an emerging set of techniques that are well suited to understanding plant-soil interactions (Michalet et al. 2013, van Dam and Bouwmeester 2016, Mhlongo et al. 2018, Swenson et al. 2018, Hu et al. 2018, Ristok et al. 2019) by directly measuring molecules with formula weights between 75 and 1000 Da involved with cellular processes; such techniques are easily adapted to probe the complex chemistry that links plants to soils. This linkage may be especially evident across gradients, such as soil nutrients or climate, which change plant

physiology and cellular activities in response to different abiotic conditions such as soil moisture (Swenson et al. 2018). Much of the current metabolomics work involving plant tissues has used simplified systems for 'proof-of-concept' experiments (van Dam and Bouwmeester 2016) that are performed on model rhizospheres, often under artificial conditions such as hydroponics. Only a few studies have examined root exudates in soils, finding patterns in plant response to herbivory, variation in the metabolome associated with root morphology, and an overall plant conditioning effect (Marti et al. 2013, Michalet et al. 2013, Pétriacq et al. 2017, Hu et al. 2018). For example, Ultra Performance Liquid Chromatography- High Resolution Mass Spectrometry (UPLC-HRMS) has been used to examine exudates from pots containing potting soil and *Arabidopsis thaliana* in a controlled greenhouse, showing that plant presence altered the extracted metabolome from the soil (Pétriacq et al. 2017). These recent successes show metabolomics may be key to examining complex biological systems, especially the rhizosphere. Moreover, Hu and coauthors (2018) show how changes in root metabolites can alter success in future plant generations showing a plant-soil feedback mediated by root metabolomics (Hu et al. 2018). Despite the many important ecological and evolutionary outcomes resulting from plant-soil interactions, no studies, to our knowledge, have explored the root metabolome and the soil organic compounds of paired root and rhizosphere soil samples to examine how population level tree trait variation can govern plant-soil signaling. This understanding should allow for greater predictive power in understanding plant-soil linkages and feedbacks under a range of ecological and evolutionary contexts. Metabolomics experiments designed to understand ecologically relevant gradients would provide much-needed detail explaining plant-soil linkages and feedbacks.

Using metabolomics and a well-known system with strong plant-soil linkages and feedbacks (*Populus angustifolia* James), we examined the genetic basis to population-level variation in the root metabolome and how the soil rhizosphere metabalome[1] was related to variation in root metabolites. *Populus angustifolia* is a model system for exploring the complexity of the rhizosphere. [1]*(Footnote: Here we use the notation soil rhizosphere metabolome, with the understanding that soil rhizosphere chemistry is a complex matrix that could encompass root exudates, microbial metabolomes and extra-cellular soil chemical processes.)* Previous experimental work has shown genetic differentiation in tree molecular and quantitative traits at the genotype-, population- and provenance-level suggesting that *P. angustifolia* has diverged in response to many biotic and

abiotic variables across its distribution (Schweitzer *et al.*, 2011; Evans *et al.*, 2014, 2015; Van Nuland *et al.*, 2017, 2018; Ware et al. 2019). Across its range, *P. angustifolia* varies in many traits that can alter belowground interactions, including plant secondary metabolites. For example, variation in polyphenol content among plant genotype explains variation in soil microbial community composition, total microbial biomass and soil extracellular enzyme activity (Madritch et al. 2009, Schweitzer et al. 2011). Genetic variation in *P. angustifolia* traits has also been shown to influence field soil conditions such as bacteria/fungal and carbon (C) to nitrogen (N) ratios (Van Nuland et al. 2017, 2018, Ware et al. 2019). To understand the genetic differences in variation between plant and soil metabolomes and if roots can influence soil chemical metabolomes, variation was explored within the root and soil rhizosphere metabolites from six unique populations of *P. angustifolia* grown in a common environment. We hypothesized and found that genetic variation among these populations, due to their evolutionary history in response to distinct environmental conditions in the field (Van Nuland et al. 2018, Ware et al. 2019), led to different root chemical phenotypes. A further hypothesis that root chemical phenotypes would have a direct impact on rhizosphere soil chemical phenotypes led to the observation that tree individuals were conditioning rhizosphere soil primarily with unique lipids, condensed hydrocarbons, and lignin.

Materials and Methods

Plant Collections

In 2012, we collected plant cuttings from six populations of narrowleaf cottonwood (*Populus angustifolia*) across the species range in the intermountain west. The six populations sampled were: the Yellowstone River (YEL, 45.636, -110.571), the Shoshone River (SHO, 43.177, -110.984), the Lexington River (LEX, 38.859, -114.208), the Dolores River (DOL, 37.595, -108.107), Park Creek River (PK, 37.679, -106.603), and the Blue River (BL, 33.698, -109.069) (Fig. 1).The cuttings of individual genotypes were collected in replicate, transported cold (~4°C) to a greenhouse at Northern Arizona University where they were allowed to establish in potting mix (equal parts peat, perlite and vermiculite). The cuttings were watered uniformly every week under ambient light conditions. After two months of growth in Arizona, the plants were transported to a greenhouse at the University of Tennessee where they grew for four years in a greenhouse before being sampled. Plants were grown in 0.5 gallon dee pots,

randomized in the greenhouse and watered bi-weekly; a 15-15-15, nitrogen-phosphorus-potassium, slow release fertilizer was applied annually.

Extraction Method for Root and Soil Metabolome Profiles

Samples of fine roots and rhizosphere soil were taken from four genotypes of each of the six populations. Unfortunately, one root sample from the Blue River and one paired root and soil sample from Park Creek were lost and could not be used in the analysis. Roots from 5 to 10 cm below soil surface were brushed clean of soil with a fine brush, before being rinsed with deionized water. The soil brushed from the roots was saved as the rhizosphere sample. Both root and rhizosphere soil samples were frozen using liquid nitrogen, then homogenized and ground to a fine powder with a mortar and pestle in a cold room (~4°C). Aliquots (~100 mg) were transferred to individual sample tubes where all subsequent extraction steps were performed. The metabolomic extraction protocol for the root and rhizosphere soil was adapted from a previously reported method (Stough et al. 2016). Specifically, 1.3 mL of a 2:2:1 methanol: acetonitrile: water with formic acid at a final concentration of 0.1M extraction solution was added and shaken on a vortex mixer for 20 minutes at 4°C. Centrifugation (17,000 g) was utilized for 5 minutes to pelletize the sample, followed by the transfer of the supernatant to a clean sample tube. The sample had 200µL of extraction solution added and the sample was subjected to the same extraction conditions previously stated. The two supernatants were combined and were evaporated using nitrogen gas. Sterile water at 4° C was used to re-suspend the dry extract before UPLC-HRMS analysis.

Ultra Performance Liquid Chromatography- High Resolution Mass Spectrometry

Separation of the metabolites was performed on a Dionex Ultimate liquid chromatography system using a Synergi 2.5µ Hydro-RP100Å, 100 mm x 2.00 mm column (Phenomenex, Torrance, CA, USA). A similar elution gradient utilizing acidic water with tributylamine and methanol over 25 minutes was employed (Lu et al. 2010) The eluent was immediately introduced to an Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) via electrospray ionization (ESI) with a capillary temperature of 300°C, spray voltage of 3 kV, nitrogen sheath and sweep gas at 25 and 3 units respectively. Data acquisition was done in negative ion mode over the range of 72 – 1000 *m/z* at 140,000 resolution with automatic gain control of $3x10^6$ ions.

Raw data files were converted with MSConvert from Proteowizard (Holman et al. 2014) then uploaded to MAVEN (Clasquin et al. 2012). Metabolites were annotated using exact mass of the $[M-H]^-$ ion and known retention times generated from an in-house curated database. Area under the curves were compiled for all samples for each metabolite simultaneously and data was normalized to the wet mass of soil or roots extracted.

Unknown Mass Spectrometry Data Reduction

Spectral features (*m/z*-retention-time pair) were identified by XCMS with a ±5 ppm error window (Tautenhahn et al. 2012), and the CAMERA package (Kuhl et al. 2012) in R (R Development Core Team, 2016) was used to identify potential isotopes and adducts. For a feature to be included in the molecular formula analysis, it must have a signal that is three times that of the blank and must be present in all the replicants in at least one of the sample groups. To avoid weighting the data towards compounds that were detected as multiple chemical species, features were removed that were annotated as the $[M+n]^-$ isotope or identified as an adduct. This reduced spectral feature dataset was used as an input to the Seven Golden Rules (Kind and Fiehn 2007) to generate potential molecular formulas. The formulas were restricted to a mass accuracy of ±5 ppm and to the following elements: carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S).

Statistical Analyses

To examine root chemical phenotypes we used a distance based redundancy analysis (dbRDA) with a Jaccard distance matrix to assign variation in the root matrix to the six populations. To address the hypothesis that root phenotypes vary by population, we tested this model against a null dbRDA model (population removed) using likelihood ratio test (α=0.05). If the population model was found to be significantly different than the null model, we interpreted the proportion of variation explained by the constrained axes in the full model as the variation attributed to differences between populations. These statistical tests were also done for the soil metabolite data. Once dbRDA models were shown to better predict the data, indicator species analysis was used to identify compounds important to the separation between groups. Coarse identification of the unknown indicator metabolites was done with the help of the chemical formulas generated from Seven Golden Rules (Kind and Fiehn 2007) and the Van Krevelen diagrams (Van Krevelen 1950, Minor et al. 2014, Brockman et al. 2018). Compounds were

sorted into classes based upon their hypothesized atomic O/C, H/C and C/N ratios. Classification of loadings of significant dbRDA models were examined to determine if the variation among populations was more likely due to differences in total production (majority of high loadings with the same sign) or tradeoffs (high loadings share positive and negative effects).

To determine if plant roots are influencing the surrounding soil, correlations between the root data and the soil data were performed by co-inertia analysis (Doledec & Chessel, 1994). A significant model (α=0.05) would suggest that the multivariate correlation coefficient (RV) can be interpreted as the relationship in shared multivariate space between a plant root conditioning the soil associated with it. Similar to a Pearson r, by squaring this value, we can interpret what amount of the total variation in soil metabolites is explained by root chemistry.

All statistical models and figures were made using R (R Development Core Team, 2016). The redundancy analysis were performed with the package vegan (Oksanen et al. 2016), the indicator species analysis were performed with the package indicspecies (Caceres & Jansen 2016) and the co-inertia was performed in the package ade4 (Dray and Dufour 2007).

Results

Over 12,000 unique root and over 5,000 unique soil organic chemical compounds were measured from plants originating from six distinct populations when growing in a common environment (Table 2.1; Ware et al., 2019). In support of the hypothesis that past population genetic divergence impacted plant root metabolomes a redundancy analysis (RDA), where canonical axes were constrained by population, showed that plant population had a significant effect on root and soil metabolomes (the six rivers sampled, north to south were: Yellowstone River MT, the Shoshone River WY, the Lexington River NV, the Dolores River CO, Park Creek River CO, and the Blue River NM, see Figure 2.1). The model including source population was better than the null model, explaining 29.8% of the variation in compounds found in plant roots (p=0.032, Table 2.1, Figure 2.2). Similarly, we found that associated rhizosphere soils also showed population-level differences, with 27.2% of the variation in the rhizosphere soil metabolome (p=0.025, Table 2.1, Figure 2.2) attributed to plant population of origin. These populations also have a hierarchical genetic structure and belong to one of three distinct genetic provenances (Figure 2.1, (Evans et al. 2015)). Redundancy analysis, however, with genetic

provenance as the two constrained axes, was not significantly better than the null model, suggesting that this level of genetic structuring does not predict differences in root ($p = 0.329$, Table 2.1) or soil metabolomes ($p=0.193$, Table 2.1).

Analysis of the RDA loadings suggests that trade-offs, and not simply abundance of metabolites, are the main drivers of the difference among populations. For example, if the abundance of metabolites is the primary driver of population-level differences in metabolomes, the majority of the loadings would be in one direction. In contrast to this result, we found many strongly positive and negative loadings on the constrained axis, suggesting unique trade-offs in allocation to chemical relative abundance (e.g., the relative abundance of unknown metabolite at mass 299.0984 was 68% different between populations where it was the most and least abundant). Indicator species analysis suggests specific metabolites are responsible for the variation among populations (Figure 2.3). Interestingly, it is only the unknown metabolites that are significantly driving the differences among populations in both the root and soil, although this result is likely due to the bias of metabolite libraries towards metabolites involved in primary metabolism and phospholipid biosynthesis. Indicator analysis does not indicate any identified metabolites. The unknown compounds are often localized in the lipid and protein regions. Park Creek River, Shoshone River, and the Blue River are populations with the greatest number of unique root metabolites. Park Creek River, however, is the only population with more unique soil metabolites than root metabolites (Figure 2.3). Despite the variation in unique unknowns, the metabolite C to N and C to hydrogen (H) ratios estimated for the root metabolites showed only a marginal trend that may suggest that Dolores River trees are producing unique compounds which have fewer H relative to the number of C atoms than the other rivers (ANOVA, C to H: $F_{(5,111)}=2.2511$, $p=0.054$). Furthermore, there was no difference between metabolite C to N ratios among populations (ANOVA, C to N: $F_{(5,111)}=1.015$, $p=0.412$) for the root metabolites. In the rhizosphere soils, there was no relationship between both metabolite C to N and C to H ratios and population.

A co-inertia analysis exploring the relationship between paired root and soil metabolome data in a shared multivariate space found support for the second hypothesis that plant cellular processes result in root metabolomes that influence soil microbial rhizosphere metabolomes. There was a significant multivariate correlation coefficient (RV) of 0.385 suggesting 15% of the

total variation in soil metabolites can be attributed to plant root metabolites (Figure 2.4). Collapsing the covariation of the root and soil metabolite matrices into two dimensions is difficult. However, Park Creek River, the Dolores River, and the Shoshone River each have distinct clusters of shared root and soil metabolites in multivariate space.

Discussion

This is the first study, to our knowledge, to explore the full metabolomic complexity of roots and rhizosphere soils from plants of the same species (*Populus angustifolia)* collected across population genetic scales. There were three important results. First, there was among population-level genetic variation for root metabolites. Significant variation among populations when grown in a common environment indicates that evolutionary processes have been a key factor in the genetic divergence among populations in root metabolome phenotypes. Second, associated soil organic chemical communities also varied by plant population and were correlated with plant root metabolomes. The complexity of plant-soil chemical interactions is often unaccounted for, despite the intimate interactions between roots and surrounding rhizosphere soil. While the root tissue metabolome sampled is a measure of not only the direct autotrophic metabolites produced, but also includes the wide variety of microbial endophytes living within and on the root tissue, root phenotypes altered soils in unique ways, leading to variation in soil metabolites among populations that initially began the greenhouse experiment in a common matrix. Multivariate analysis reveals some shared structure to the paired root and soil datasets after only 4 years of plant growth, showing that plant roots have strong impacts on soil metabolite profiles. Third, these results indicate that there is a wide variety of compounds relating above- and belowground linkage and feedback between plants and soils. In fact, because most of the variation in plant-soil metabolomes are in unknown compounds, it may be important to revise both the compound libraries (i.e., within XCMS) as well as chemical metabolic pathways in plants that may be important in linking plant and soil function.

Evolution of Rhizosphere Chemical Composition and Plant-Soil Conditioning

Evolution of the root metabolome among populations of *P. angustifolia* suggests that plants have diverged in important physiological/metabolic processes across the landscape. Importantly, the RDA loadings suggests that variation among populations is not driven in a single direction, but that tradeoffs between compounds describe the patterns. Indeed, over 29%

of the variation in root metabolome within *P. angustifolia* is attributed to population-level variation when plants were grown in a common environment. Interestingly, genetic provenance was not distinguished as a significant predictor of root tissue and rhizosphere soil metabolomes. This may have been due to the low replication at the provenance level. To properly examine variation due to provenance, populations are the necessary experimental unit, and six populations spread across three groups may have had insufficient power to detect a significant effect. However, at the population scale, these result suggests that there are multiple pathways by which plants alter soil microbes and plant-microbe interactions, potentially driving diverse belowground microbial function. For example, plant chemical functions have been shown to vary within *P. angustifolia*, and alter plant-soil linkages (Whitham et al. 2006, Schweitzer et al. 2011, 2014). Broadly, plant root exudates have been linked to signaling of beneficial microbes (Kiers and Denison 2008), direct defense (Baetz and Martinoia 2014), signaling between plants (Chamberlain et al. 2001) as well as allelochemicals reducing competition (Hu et al. 2018). Because the root metabolome is the plant trait that most intimately interacts with the soil, probing variation of the rhizosphere organic composition provides a more complete picture of plant chemical profiles and allows analysis of the consequences of these differences.

Among population-level differences in root metabolomes altered soils in unique ways, resulting in variation to soil metabolites among populations that initially began the experiment in a common potting mix. There were population specific correlations between over 12,000 unique metabolites in plant roots and over 5000 unique metabolites in rhizosphere soils. Increasing evidence indicates that plant genetics can predictably affect associated community structure, even in soil microbes (Schweitzer et al. 2008, Madritch et al. 2009, Schweitzer et al. 2011, Hu et al. 2018). However, we believe these are the first data to demonstrate that the soil chemical community responds differently to different populations of plants. Finding the 15% effect of plant root metabolome on the associated soil is a deceptively small finding. With the total variation in metabolites also altered by factors such as soil moisture content, aggregate size, temperature, time of day, time of sampling, and many other sources of variation, it is surprising that plant roots show any linkage in these data.

No Single Chemical Pathway Links Plants and Soils

Despite incomplete knowledge of the chemical structure of all the compounds detected, this study is an important advancement in our understanding of plant-soil interactions. It has been hypothesized that plants make hundreds of thousands of metabolites (Dixon and Strack 2003, Yonekura-Sakakibara and Saito 2009, Pichersky and Lewinsohn 2011). Within our root tissues, we identified over 12,000 unique metabolites. The chemical richness revealed and examined in this studies 10 times greater than the chemical richness reported in the other studies examining plants and soils in tandem (Pétriacq et al. 2017). Visualizing the data among populations with Van Krevelen diagrams allows for characterization of the unidentified chemical species into probable classes of metabolite based on chemical formulae (Minor et al. 2014, Brockman et al. 2018), and the bulk of the detected unknown metabolites were classified as lipids, condensed hydrocarbons, and lignin. These diagrams only show compounds unique to a population, and are not suggesting that other compounds are not important to plant-soil conditioning, only that they are not uniquely high or low in abundance across the populations. Unique compounds among populations in both root and soil samples suggest that the variation in rhizosphere metabolomes between populations is, in part due to variation in plant root chemistry that varies across populations.

The scope of the findings suggests a need to increase efforts to build sample libraries for complex metabolites. While options for interpreting unknown compounds exist (Brockman et al. 2018) improving the currently available sample libraries is critical (Mhlongo et al. 2018). Despite the current lack of comprehensive libraries, and the common limitations of analytical techniques (e.g., solvent choice and detection range) both of which influence the classes of compounds that are detected and then interpreted, these results show that metabolite patterns can be detected and analyzed across landscapes.

Conclusions & Future Directions

Researchers have been explicitly examining plant-soil interactions and feedbacks for over two decades (Bever et al. 1997, van der Putten 1997), and indirectly for centuries since humans started cultivating crops. Over time, understanding of these interactions have been improved with the addition of plant genetics, soil structure and chemical properties, rhizosphere and root chemistry, and soil microbial genetics (Badri et al. 2009, van Dam and Bouwmeester 2016,

Rasmann and Turlings 2016). Our findings indicate the importance of ecologically and evolutionarily meaningful plant-soil experiments with metabolomics. What started in hydro/aeroponic systems to examine root metabolomes (Khorassani et al. 2011, Jandová et al. 2015) has moved into sterile sands, glass beads, and potting mixes (Marti et al. 2013, Escudero et al. 2015, Pétriacq et al. 2017). Many studies to date ignore genetic variation within target plant species, with a few key exceptions where genetic variation in the root metabolome has been explicitly tested and found to differ between two species of *Populus* (Tschaplinski et al. 2014) or specific plant gene manipulations (Hu et al. 2018). Furthermore, while examining pot-level exudate metabolomes is an important method (Pétriacq et al. 2017), future studies should be willing to explore explicit components of the rhizosphere (e.g., separating soil aggregates, root ages). Interpreting the consequences of invasion, range shift, or succession without considering the complexity of the rhizosphere eliminates a key interaction driving community dynamics. This study has shown that the chemical complexity of plant-soil interactions is much greater than often reported, yet can be explored and interpreted. Further, there is a large amount of variation in rhizosphere chemistry due to population-level divergence across landscapes. As we begin to build predictive models of species migration that incorporate species interactions and ecosystem change (Harsch et al. 2017), it is critical to understand the underlying mechanisms of these plant-soil interactions. Together, these results show that metabolomics and chemical profiling is a reduced bias means to identify and interpret the chemical complexity of the rhizosphere to understand the mechanistic linkages between plant physiological processes and soil chemistry. Understanding the role of evolutionary divergence in plant-microbial-soil interactions may be key to future understanding of a range of processes, including plant community succession, invasion success, patterns of biodiversity and aspects of ecosystem function in a changing world.

Acknowledgments

I would like to thank Ian Ware, Michael van Nuland, Philip Patterson, and Courtney Gorman for assistance in the field and greenhouse as well as Terrell Carter and Michaela Humby for their help in the lab. Thank you to Melissa Liotta and Shannon Bayliss for their help building figures. Thanks to Ken McFarland and Jeff Martin for their greenhouse expertise. I would like to acknowledge funding from The University of Tennessee, Department of Ecology and Evolutionary Biology.

References

Badri, D. V., T. L. Weir, D. van der Lelie, and J. M. Vivanco. 2009. Rhizosphere chemical dialogues: plant–microbe interactions. Current Opinion in Biotechnology 20:642–650.

Baetz, U., and E. Martinoia. 2014. Root exudates: the hidden part of plant defense. Trends in Plant Science 19:90–98.

Bardgett, R. D., and D. A. Wardle. 2010. Aboveground-Belowground Linkages Biotic Interactions, Ecosystem Processes, and Global Change. Oxford University Press, Oxford.

Bever, J. D., K. M. Westover, and J. Antonovics. 1997. Incorporating the Soil Community into Plant Population Dynamics: The Utility of the Feedback Approach. The Journal of Ecology 85:561.

Bezemer, T. M., C. S. Lawson, K. Hedlund, A. R. Edwards, A. J. Brook, J. M. Igual, S. R. Mortimer, and W. H. Van Der Putten. 2006. Plant species and functional group effects on abiotic and microbial soil properties and plant-soil feedback responses in two grasslands. Journal of Ecology 94:893–904.

Brockman, S. A., E. V. Roden, and A. D. Hegeman. 2018. Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. Metabolomics 14:48.

Chamberlain, K., E. Guerrieri, F. Pennacchio, J. Pettersson, J. A. Pickett, G. M. Poppy, W. Powell, L. J. Wadhams, and C. M. Woodcock. 2001. Can aphid-induced plant signals be transmitted aerially and through the rhizosphere? Biochemical Systematics and Ecology 29:1063–1074.

Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Current protocols in bioinformatics Chapter 14:Unit14.11.

Coskun, D., D. T. Britto, W. Shi, and H. J. Kronzucker. 2017. How Plant Root Exudates Shape the Nitrogen Cycle 22:661673

Crutsinger, G. M., N. J. Sanders, and A. T. Classen. 2009. Comparing intra- and inter-specific

effects on litter decomposition in an old-field ecosystem. Basic and Applied Ecology 10:535–543.

van Dam, N. M., and H. J. Bouwmeester. 2016. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. Trends in Plant Science 21:256-265

Dixon, R. A., and D. Strack. 2003. Phytochemistry meets genome analysis, and beyond..... Phytochemistry 62:815–816.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species–environment relationships. Freshwater Biology 31:277–294.

Dray, S., and A.-B. Dufour. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software 22.

Escudero, N., F. C. Marhuenda-Egea, R. Ibanco-Canête, E. A. Zavala-Gonzalez, and L. V. Lopez-Llorca. 2015. A metabolomic approach to study the rhizodeposition in the tritrophic interaction: Tomato, Pochonia chlamydosporia and Meloidogyne javanica. Metabolomics 10:788–804.

Evans, L. M., G. J. Allan, S. P. Difazio, G. T. Slavov, J. A. Wilder, K. D. Floate, S. B. Rood, and T. G. Whitham. 2015. Geographical barriers and climate influence demographic history in narrowleaf cottonwoods. Heredity 114:387–396.

Evans, L. M., G. T. Slavov, E. Rodgers-Melnick, J. Martin, P. Ranjan, W. Muchero, A. M. Brunner, W. Schackwitz, L. Gunter, J. G. Chen, G. A. Tuskan, and S. P. Difazio. 2014. Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations. Nature Genetics 46:1089–1096.

Harsch, M. A., A. Phillips, Y. Zhou, M.-R. Leung, D. S. Rinnan, and M. Kot. 2017. Moving forward: insights and applications of moving-habitat models for climate change ecology. Journal of Ecology 105:1169–1181.

Hobbie, S. E., P. B. Reich, J. Oleksyn, M. Ogdahl, R. Zytkowiak, C. Hale, and P. Karolewski. 2006. Tree species effects on decomposition and forest floor dynamics in a common garden. Ecology 87:2288–2297.

Holman, J. D., D. L. Tabb, and P. Mallick. 2014. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. Pages 13.24.1-13.24.9 Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Hu, L., C. A. M. Robert, S. Cadot, X. Zhang, M. Ye, B. Li, D. Manzo, N. Chervet, T. Steinger, M. G. A. van der Heijden, K. Schlaeppi, and M. Erb. 2018. Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. Nature Communications 9:2738.

Jandová, K., P. Dostál, T. Cajthaml, and Z. Kameník. 2015. Intraspecific variability in allelopathy of Heracleum mantegazzianum is linked to the metabolic profile of root exudates. Annals of Botany 115:821–831.

Kardol, P., T. Martijn Bezemer, and W. H. Van Der Putten. 2006. Temporal variation in plant-soil feedback controls succession. Ecology Letters 9:1080–1088.

Khorassani, R., U. Hettwer, A. Ratzinger, B. Steingrobe, P. Karlovsky, and N. Claassen. 2011. Citramalic acid and salicylic acid in sugar beet root exudates solubilize soil phosphorus. BMC Plant Biology 11:121.

Kiers, E. T., and R. F. Denison. 2008. Sanctions, Cooperation, and the Stability of Plant-Rhizosphere Mutualisms. Annual Review of Ecology, Evolution, and Systematics 39:215–236.

Kind, T., and O. Fiehn. 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8:105.

Van Krevelen, D. 1950. Graphical statistical method for the study of structure and reaction processes of coal. Fuel 29:269–284.

Kuhl, C., R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. 2012. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry 84:283–289.

Lojewski, N. R., D. G. Fischer, J. K. Bailey, J. A. Schweitzer, T. G. Whitham, and S. C. Hart. 2012. Genetic components to belowground carbon fluxes in a riparian forest ecosystem: A common garden approach. New Phytologist 195:631–639.

Lu, W., M. F. Clasquin, E. Melamud, D. Amador-Noguez, A. A. Caudy, and J. D. Rabinowitz. 2010. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Analytical Chemistry 82:3212–3221.

Madritch, M. D., S. L. Greene, and R. L. Lindroth. 2009. Genetic mosaics of ecosystem functioning across aspen-dominated landscapes. Oecologia 160:119–127.

Madritch, M. D., and M. D. Hunter. 2005. Phenotypic variation in oak litter influences short- and long-term nutrient cycling through litter chemistry. Soil Biology and Biochemistry 37:319–327.

Marti, G., M. Erb, J. Boccard, G. Glauser, G. R. Doyen, N. Villard, C. A. M. Robert, T. C. J. Turlings, S. Rudaz, and J. L. Wolfender. 2013. Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. Plant, Cell and Environment 36:621–639.

Melillo, J. M., J. D. Aber, and J. F. Muratore. 1982. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. Ecology 63:621–626.

Mhlongo, M. I., L. A. Piater, N. E. Madala, N. Labuschagne, and I. A. Dubery. 2018. The Chemistry of Plant–Microbe Interactions in the Rhizosphere and the Potential for Metabolomics to Reveal Signaling Related to Defense Priming and Induced Systemic Resistance. Frontiers in Plant Science 9:112.

Michalet, S., J. Rohr, D. Warshan, C. Bardon, J. C. Roggy, A. M. Domenach, S. Czarnes, T. Pommier, B. Combourieu, N. Guillaumaud, F. Bellvert, G. Comte, and F. Poly. 2013. Phytochemical analysis of mature tree root exudates in situ and their role in shaping soil microbial communities in relation to tree N-acquisition strategy. Plant Physiology and Biochemistry 72:169–177.

Minor, E. C., M. M. Swenson, B. M. Mattson, and A. R. Oyler. 2014. Structural characterization of dissolved organic matter: A review of current techniques for isolation and analysis.

Van Nuland, M. E., J. K. Bailey, and J. A. Schweitzer. 2017. Divergent plant–soil feedbacks could alter future elevation ranges and ecosystem dynamics. Nature Ecology & Evolution 1:0150.

Van Nuland, M. E., I. M. Ware, J. K. Bailey, and J. A. Schweitzer. 2018. Geographic variation in plant evolutionary responses to soil fertility is mediated by soil microbial communities. Functional Ecology 33:95–106.

Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2016. Vegan: Community Ecology Package. https://github.com/vegandevs/vegan.

Pétriacq, P., A. Williams, A. Cotton, A. E. McFarlane, S. A. Rolfe, and J. Ton. 2017. Metabolite profiling of non-sterile rhizosphere soil. Plant Journal 92:147–162.

Pichersky, E., and E. Lewinsohn. 2011. Convergent Evolution in Plant Specialized Metabolism. Annual Review of Plant Biology 62:549–566.

Pregitzer, C. C., J. K. Bailey, and J. A. Schweitzer. 2013. Genetic by environment interactions affect plant-soil linkages. Ecology and Evolution 3:2322–2333.

van der Putten, W. H. 1997. Plant-soil feedback as a selective force. Trends in Ecology & Evolution 12:169–170.

van der Putten, W. H. 2003. Plant defense belowground and spatiotemporal processes in natural vegetation. Ecology 84:2269–2280.

van der Putten, W. H., M. A. Bradford, E. Pernilla Brinkman, T. F. J. van de Voorde, and G. F. Veen. 2016. Where, when and how plant–soil feedback matters in a changing world. Functional Ecology 30:1109–1121.

Rasmann, S., and T. C. J. Turlings. 2016. Root signals that mediate mutualistic interactions in the rhizosphere. Current Opinion in Plant Biology 32:62–68.

Ristok, C., Y. Poeschl, J.-H. Dudenhöffer, A. Ebeling, N. Eisenhauer, F. Vergara, C. Wagg, N. M. van Dam, and A. Weinhold. 2019. Plant species richness elicits changes in the metabolome of grassland species via soil biotic legacy. Journal of Ecology Early view. https://doi.org/10.1111/1365-2745.13185

Schweitzer, J. A., D. G. Fischer, B. J. Rehill, S. C. Wooley, S. A. Woolbright, R. L. Lindroth, T. G. Whitham, D. R. Zak, and S. C. Hart. 2011. Forest gene diversity is correlated with the

composition and function of soil microbial communities. Population Ecology 53:35–46.

Schweitzer, J. A., I. Juric, T. F. J. van de Voorde, K. Clay, W. H. van der Putten, and J. K. Bailey. 2014. Are there evolutionary consequences of plant-soil feedbacks along soil gradients? Functional Ecology 28:55–64.

Schweitzer, J. A., M. D. Madritch, J. K. Bailey, C. J. Leroy, D. G. Fischer, B. J. Rehill, R. L. Lindroth, A. E. Hagerman, S. C. Wooley, S. C. Hart, and T. G. Whitham. 2008. From genes to ecosystems: The genetic basis of condensed tannins and their role in nutrient regulation in a Populus model system. Ecosystems 11:1005-1020

Stough, J. M. A., S. P. Dearth, J. E. Denny, G. R. LeCleir, N. W. Schmidt, S. R. Campagna, and S. W. Wilhelm. 2016. Functional Characteristics of the Gut Microbiome in C57BL/6 Mice Differentially Susceptible to Plasmodium yoelii. Frontiers in Microbiology 7:1520.

Swenson, T. L., U. Karaoz, J. M. Swenson, B. P. Bowen, and T. R. Northen. 2018. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. Nature Communications 9:19.

Tautenhahn, R., G. J. Patti, D. Rinehart, and G. Siuzdak. 2012. XCMS online: A web-based platform to process untargeted metabolomic data. Analytical Chemistry 84:5035–5039.

Team, R. D. C., and R. R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Tschaplinski, T. J., J. M. Plett, N. L. Engle, A. Deveau, K. C. Cushman, M. Z. Martin, M. J. Doktycz, G. A. Tuskan, A. Brun, A. Kohler, and F. Martin. 2014. Populus trichocarpa and Populus deltoides Exhibit Different Metabolomic Responses to Colonization by the Symbiotic Fungus Laccaria bicolor. Molecular Plant-Microbe Interactions 27:546–556.

Vitousek, P. M., L. R. Walker, L. D. Whiteaker, D. Mueller-Dombois, and P. A. Matson. 1987. Biological Invasion by Myrica faya Alters Ecosystem Development in Hawaii. Science 238:802–804.

Ware, I. M., M. E. Van Nuland, J. A. Schweitzer, Z. Yang, C. W. Schadt, L. C. Sidak-Loftis, N. E. Stone, J. D. Busch, D. M. Wagner, and J. K. Bailey. 2019. Climate-driven reduction of genetic variation in plant phenology alters soil communities and nutrient pools. Global

Change Biology 25:1514–1528.

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. LeRoy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, B. M. Potts, D. G. Fischer, C. A. Gehring, R. L. Lindroth, J. C. Marks, S. C. Hart, G. M. Wimp, and S. C. Wooley. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics 7:510–523.

Wolfe, B. E., V. L. Rodgers, K. A. Stinson, and A. Pringle. 2008. The invasive plant Alliaria petiolata (garlic mustard) inhibits ectomycorrhizal fungi in its introduced range. Journal of Ecology 96:777–783.

Yonekura-Sakakibara, K., and K. Saito. 2009. Functional genomics for plant natural product biosynthesis. Natural Product Reports 26:1466.

Appendix

Table 2.1: Results of the redundancy analysis by population and provenance. Redundancy analysis performed on the root metabolome data and the soil metabolome data by both *Populus angustifolia* population (n=6) and provenance (n=3).

	Number of Metabolites	Proportion of variance explained					Comparison to null model
		RD1	RD2	RD3	RD4	RD5	
Roots	12,884						
By Population		0.1164	0.0673	0.0482	0.0444	0.0213	p = 0.032
By Provenance		0.0500	*NA*	*NA*	*NA*	*NA*	p = 0.329
Soils	5441						
By Population		0.0863	0.0703	0.0594	0.0309	0.0254	p = 0.025
By Provenance		0.0549	*NA*	*NA*	*NA*	*NA*	p = 0.193

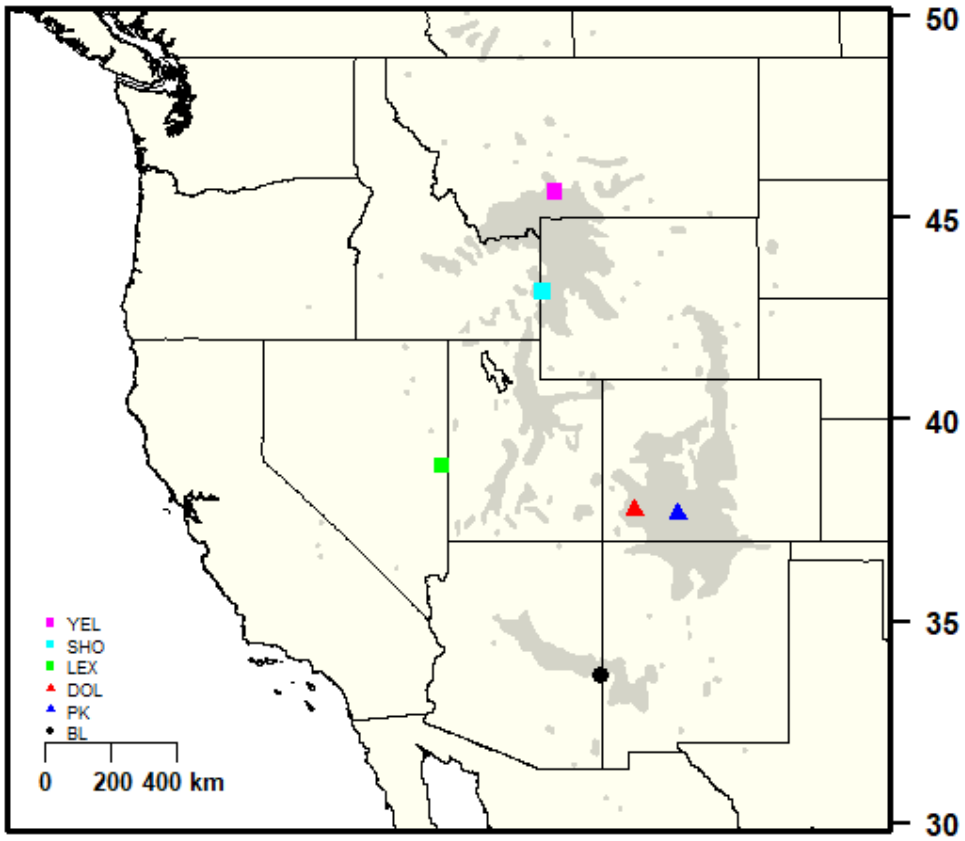


Figure 2.1: Sample distribution and sampling sites. A map showing the distribution of the six rivers across the mountain west. The six populations sampled (from north to south) were: the Yellowstone River (YEL, 45.636, -110.571), the Shoshone River (SHO, 43.177, -110.984), the Lexington River (LEX, 38.859, -114.208), the Dolores River (DOL, 37.595, -108.107), Park Creek River (PK, 37.679, -106.603), and the Blue River (BL, 33.698, -109.069). These populations organize into three larger genetic groupings, with the Yellowstone River, the Shoshone River, and the Lexington River belonging to a distinct provenance (Squares on the map). The Dolores River and the Park Creek River belong to a provenance east of the Rocky Mountains (Triangles), and the Blue River to the southernmost provenance (Circle). The grey outline represents the species range.

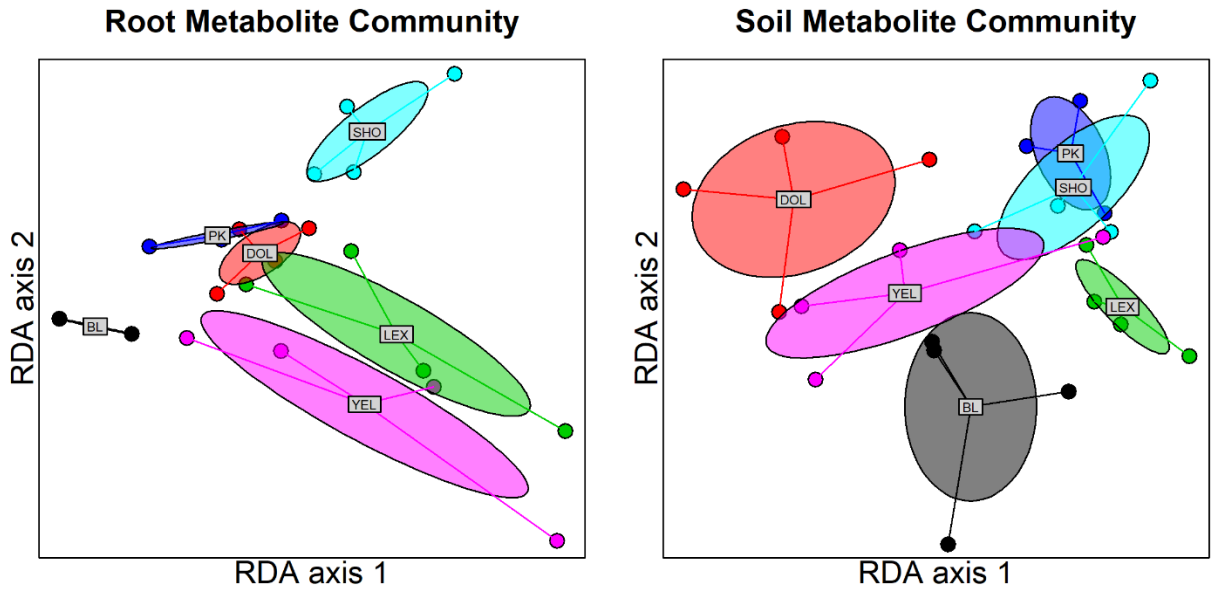


Figure 2.2: Plant populations influence root and soil metabolite communities. The first two RDA axis showing Root (left panel) and rhizosphere soil (right panel) metabolites from six rivers: The Yellowstone (YEL), The Shoshone (SHO), The Lexington (LEX), The Dolores (DOL), Park Creek (PK), and The Blue (BL). RDA analysis with population defining the constrained axes were better than the null models, explaining 30% of the variation in compounds found in roots (p=0.032) and 27% of the variation in compounds found in the rhizosphere soils (p=0.011). The further two populations are from each other represent how different their metabolite communities are from each other. Shaded ellipses represent confidence intervals around each population centroid. Filled circles are each genotype's position in multivariate space.

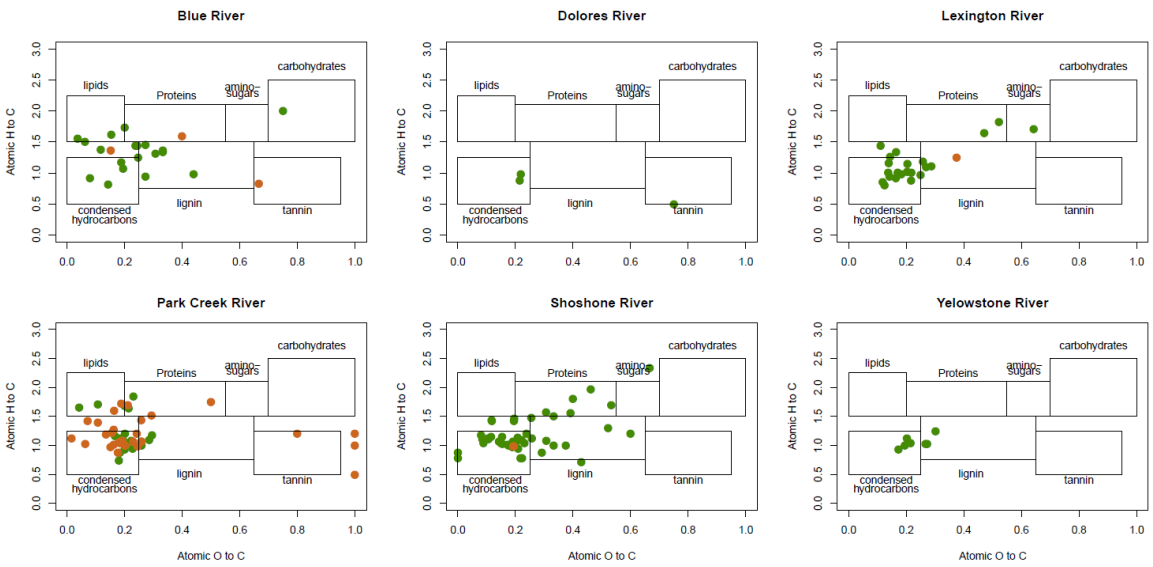


Figure 2.3: Indicator compounds for soil and roots show which families of compounds drive the population variation. Six van Krevelen diagrams showing atomic oxygen (O) to carbon (C) ratios along the x axis and atomic hydrogen (H) to carbon (C) ratios along the y axis. Points in green represent the best molecular formula estimate for a significant indicator compound differentiating that population's root metabolome from the others. Points in orange represent the best molecular formula estimate for a significant indicator compound differentiating that population's conditioned soil from the other populations. Boxes estimating the range of chemical classes were sourced from Elizabeth Minor and co-authors (Minor et al., 2014).

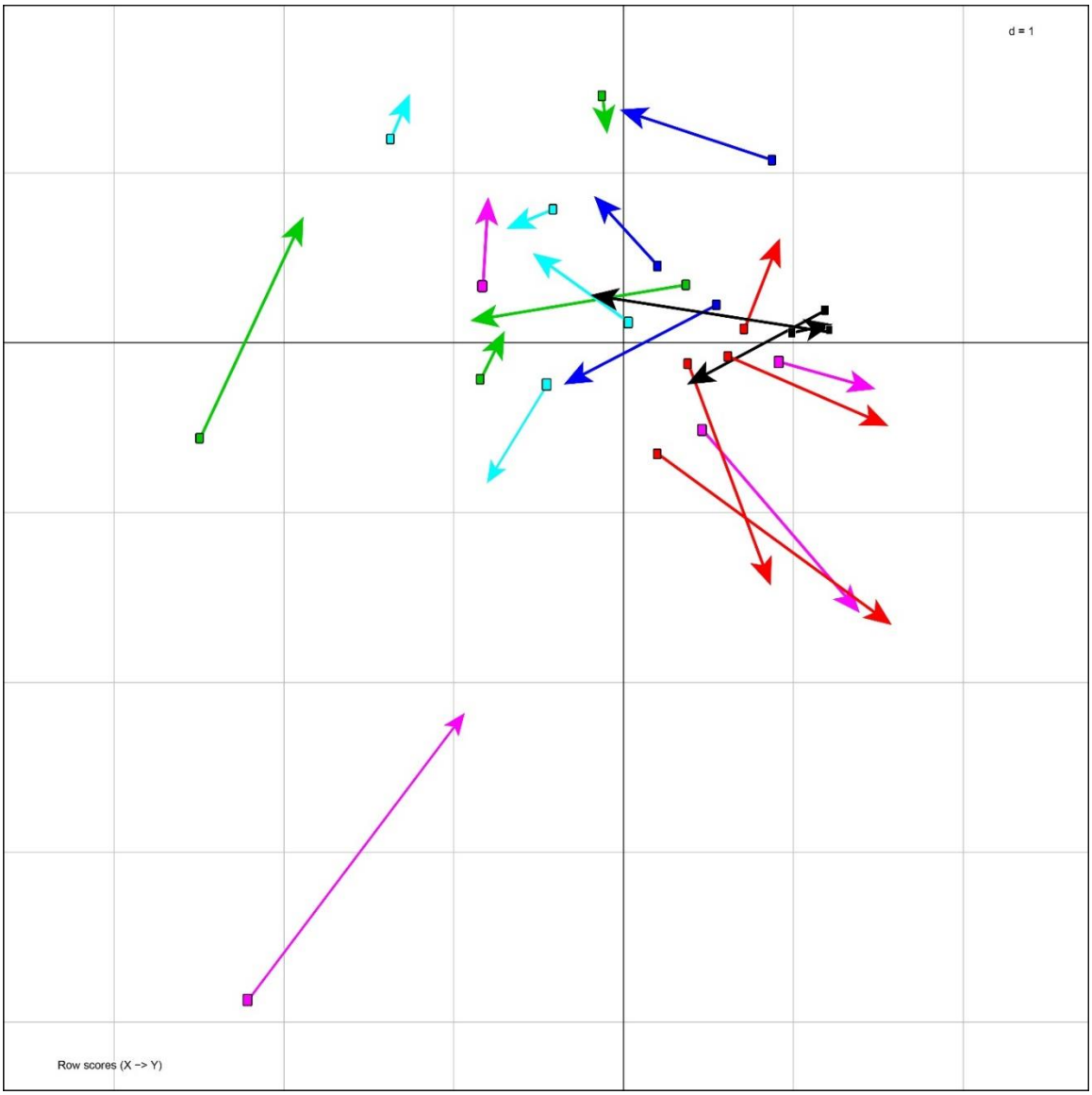


Figure 2.4: Variation in soil metabolome is correlated with variation in root metabolome. Multivariate axes 1 and 2 showing the relationship in shared multivariate space between root metabolites (base of arrow) and their paired soil metabolites (head of arrow) of each genotype. Distance between the head and tail is relative to the strength of the paired plant-soil relationship (smaller arrows are more tightly linked plant-soil signal). Distance between arrows is relative to the similarity between samples. Arrows closer together represent plants and soils that share similar root and soil metabolomes. Arrow color represents plant population (Black: Blue River, Red: Dolores River, Green: Lexington, Blue: Park Creek River, Cyan: Shoshone River, Magenta: Yellowstone River).

CHAPTER III

PLANT POPULATION INTERACTS WITH SOIL MICROBIOMES TO CREATE UNIQUE ROOT METABOLOMES

Abstract

Plants condition their surrounding soil through both above- and belowground inputs. Soil pathogens, beneficial fungi and bacteria, and root herbivores are all components of the soil with which plant roots have to successfully interact in a way that ensures plant health and fitness. To do this, plants produce a variety of compounds, from sugar exudates, to defense compounds and even fungal growth hormones that mediate the soil microbial community. In stable ecosystems, these plant-soil interactions developed over multiple generations of plant populations and soil communities. To better understand the chemical mechanisms involved in plant-soil interactions, we used four distinct populations of *Populus angustifolia* and associated soil communities in a fully factorial soil greenhouse inoculation experiment to examine how rhizosphere metabolites change in different tree-soil pairs. We hypothesized and found that the interaction between tree population and soil inoculation predicted 60% of the variation in the 9,000 root tissue metabolites measured. In the surrounding rhizosphere soil, no single model predicted the variation in the 5,000 soil metabolites. However, smaller tests on subsets of the data by tree population were different from null models, further demonstrating the strength that autotrophs have over rhizosphere chemistry. These results highlight the importance of metabolomics approaches for ecologists, acknowledge the chemical complexity of the rhizosphere and show how metabolomics can reveal multiple mechanisms governing the interactions between plants and soils.

Introduction

At a broad range of scales, the interaction of plants and soils alter both plant phenotypes and soil characteristics. Variation in soil physical, chemical and biotic characteristics have been shown to define plant distributions, fitness and a range of phenotypes, including phenology (Panke-Buisse et al. 2017, Ware et al. 2019b), plant growth, nutrient utilization (Harrison et al. 2006, Wooliver et al. 2018), and survival (Pregitzer et al. 2010, Lau and Lennon 2012). For example, plant evolution across soil gradients is common in serpentine soils, which are hotspots of plant diversity that often contain more trait variation than surrounding locations (Brady et al. 2005, Harrison et al. 2006). Furthermore, the evolution of locally adapted plant phenotypes is often found across less extreme soil gradients where available resources differ (Chapin et al. 1993, Treseder and Vitousek 2001). For example, across a gradient of soil nitrogen in Hawai`i, *Meterosideros polymorpha* showed distinct genetic separation among sites along with variation in traits associated with nitrogen cycling (Treseder and Vitousek 2001), growth, and leaf pubescence (Mueller et al. 2017). At the same time, plant traits often affect a range of edaphic characteristics, for example, soil carbon and respiration change associated with specific plant growth and chemical traits (Orwin et al. 2010, Metcalfe et al. 2011, Lojewski et al. 2012, Pregitzer et al. 2013). These individual interactions between plants and soils scale up to communities and ecosystems (Van Nuland et al. 2016, Ware et al. 2019a) whereby plant population genetics may structure belowground communities which can have ecological and evolutionary consequences for plant populations, community dynamics, and ecosystem function (Vitousek et al. 1987, Whitham et al. 2006, Bardgett and Wardle 2010). Moreover, plant conditioning of soil can facilitate plant range expansion (Bezemer et al. 2006, van der Putten et al. 2016, Van Nuland et al. 2017), improve invasion success (Van der Putten 2003, Wolfe et al. 2008), and mediate plant succession (Kardol et al. 2006). The reciprocal feedback between plants and soils, therefore, is critical for understanding the distribution of ecosystem services in terrestrial ecosystems, even if the specific mechanisms for plant-soil linkages and feedbacks is often unknown.

Plants alter their associated soil, in part, through a large variety of root exudates that can shape associated soil communities (Rovira 1969, Grayston et al. 1997). A host of plant molecular signals and receptors have been identified (Bais et al. 2006) and work that began in allelopathy

(Bertin et al. 2003) has expanded to examine the many ways plant exudates shape the rhizosphere (i.e., root associated soil) microbial community composition and function (Bezemer and van Dam 2005, Bais et al. 2006, Badri and Vivanco 2009, Dennis et al. 2010, van Dam and Bouwmeester 2016, Zhalnina et al. 2018). For example, plants alter the relative concentrations of glucose, fructose, and maltose in their roots throughout their lifetime, and variation in sugar concentration can alter the rate of bacterial colonization (Lugtenberg et al. 1999, Chaparro et al. 2013). Over a plant's lifespan the changes in root exudates (both sugars and other secondary metabolites) alter microbial traits in association with the plant (Zhalnina et al. 2018). This variation over time, and variation in root exudates of closely related individuals is enough to dramatically alter the composition of the rhizosphere microbial community (Badri et al. 2009, Mao et al. 2014, Zhalnina et al. 2018). For example, rhizosphere microbial traits, such as the metabolism of different carbohydrates and amino-acids, were shown to track changes in root exudates over time (Chaparro et al. 2013). Variation in root chemical phenotype has ecosystem level consequences.

Reciprocally, variation in rhizosphere microbial communities has been shown to influence a wide range of plant traits. Aboveground traits such as growth rate and phenology have been shown to respond to changes in soil microbial communities (Bardgett and Wardle 2010, Panke-Buisse et al. 2015, Souza et al. 2015, Ware et al. 2019b) and belowground physical traits such as root structure have also been shown to change with changes in soil microbial composition (Kudoyarova et al. 2015, Rellán-Álvarez et al. 2016). Careful multi-generational experiments have demonstrated the control that soil microbes have over plant fitness (Panke-Buisse et al. 2015, 2017) and survival (Lau and Lennon 2012). Less understood is the effect that soil microbes have on root exudate chemistry (Wurst et al. 2010, Baetz and Martinoia 2014, van Dam and Bouwmeester 2016, Chagas et al. 2018). For example, specific soil fungal pathogens such as *Fusarium* spp. and *Phythophora* spp. have been shown to induce secondary compounds in root tissues (Wurst et al. 2010), but these belowground interactions are understudied, especially in the full context of plant-soil feedbacks (Martijn Bezemer et al. 2013).

The increase in plant-soil feedback experiments (reviewed in Hendry, 2019) has been critical to understanding a broad range of landscape-level above- and belowground community processes, but there is little research examining the traits that most intimately link plants with soils, the root tissue metabolome (Hu et al., 2018, Ristok et al., 2019, Mueller et al. *in review*).

Metabolomics is an emerging technique that is well suited to understanding plant-soil interactions (Michalet et al. 2013, van Dam and Bouwmeester 2016, Mhlongo et al. 2018, Swenson et al. 2018, Hu et al. 2018) by directly measuring molecules with formula weights between 75 and 1000 Da involved with cellular processes; such techniques are easily adapted to probe the complex chemistry that link plants to soils. This linkage may be especially evident across gradients, such as soil nutrients or climate, which change plant physiology and cellular activities in response to different abiotic conditions, such as soil moisture (Swenson et al., 2018, Mueller et al., *in review*). Much of the current metabolomics work involving plant tissues has used highly simplified systems for 'proof-of-concept' experiments (van Dam and Bouwmeester 2016), often under artificial conditions such as hydroponics. Only a few studies, to date, have examined root exudates in soils, finding patterns in plant response to herbivory, variation in the metabolome associated with root morphology, and an overall plant conditioning effect (Marti et al., 2013, Michalet et al., 2013, Pétriacq et al., 2017, Hu et al., 2018, Ristok et al. 2019, Mueller et al., *in review*). For example, Ultra Performance Liquid Chromatography- High Resolution Mass Spectrometry (UPLC-HRMS) has been used to examine root tissue from pots containing soil previously conditioned by different species, showing the effect of soil inoculation on plant metabolites (Ristok et al. 2019). These recent successes show metabolomics may be key to examining complex above- and belowground biological systems, linked through the rhizosphere. Within-species effects have also been examined, Hu and co-authors (2018) show how changes in soil metabolites caused by individual plant genotypes can alter the success of future plant generations, showing plant-soil feedback mediated by root metabolomics (Hu et al. 2018). Despite the many important ecological and evolutionary outcomes resulting from plant-soil interactions, no study to our knowledge, has explored paired root and rhizosphere metabolomes to examine how the strength of plant genetic by environment (G x E) interactions may govern plant-soil interactions and their feedbacks.

Using newly developed metabolomics techniques for root and soil sampling (Mueller et al., *in review*) we explore the chemical complexity of the metabolome of root tissues and surrounding rhizosphere soils to determine how plant-soil feedbacks alter the metabolite communities expressed in the rhizosphere. These new metabolite methods have been used to identify how plant population-level genetic variation is a mechanism of plant conditioning of soils in a homogenous soil environment (Mueller et al., *in review*). The next frontier of this

research, which is examined here, is to include live soil microbial inocula to examine how plant traits and soil microbial environments interact to shape the rhizosphere metabolome. Specifically, we use four populations of *Populus angustifolia* (James), and reciprocal soil inoculations from these same populations to set up a fully factorial plant-soil inoculation experiment to address the following two hypotheses: 1) The root tissue metabolome is a function of the interaction between both the plant population and the soil environment it is exposed to; and 2) Soil rhizosphere metabolomes vary by both the soil microbial environment and in response to the tree root metabolome. A significant interaction between plant root metabolomes and field soil inoculum source would be the first demonstration of the effects of a plant-soil feedback on paired root tissue and rhizosphere soil metabolite communities. The identification of which is a frontier in environmental metabolomics (van Dam and Bouwmeester 2016)

Methods

Plant and Soil Sample Collection

In June 2017, we collected plant cuttings from four populations of narrowleaf cottonwood (*Populus angustifolia*) across the species range in the intermountain west. The four populations sampled were: Snake River (SNR, 43.70559, -110.675), Weber River (WR, 40.73897, -111.242), Oak Creek (OC, 35.06071, -111.719), and Blue River (BL, 33.698, -109.069) (Figure S3.1). Cuttings of three individual genotypes were collected in replicate at each site, transported cold (~4°C) to a greenhouse at Northern Arizona University where they were allowed to establish in potting mix (equal parts peat, perlite and vermiculite) with a 0.8% Indole-3-butyric acid (IBA) rooting powder (Hormodin 3, OHP, Inc. Bluffton, SC, USA). The cuttings were watered uniformly every week under ambient light conditions. After two months of growth in Arizona, the plants were transported to a greenhouse at the University of Tennessee where they were grown for two months before soil treatments were applied (October 2017); plants were grown at UT for nine months before the root and soil metabolomes were sampled. Plants were grown in 0.5 gallon dee pots (~2L), randomized in the greenhouse and watered bi-weekly. To combat greenhouse pests, a single application of 1% Imidacloprid,1-[(6-Ch lo ro-3-pyridinyl)methyl]-N-nitro- 2-imidazolidinimine (Marathon 1% G, OHP, Inc. Bluffton, SC, USA) was given to each plant four months prior to sampling.

Bulk soil was used as inoculation to test the interactive effects of plant genetic variation and soil communities on root and soil metabolomes. Bulk soil was collected from within 10 cm of the trunk of each of the same field trees that were collected for plant cuttings, to a depth between 0 to ~15 cm below the soil surface. Roughly 0.5 L of soil was collected from under each sampled tree. Soils were kept cold (~4° C), but not frozen, until the time of inoculation. Tree pots were reciprocally inoculated with 0.015 L of field soil (0.75 % of pot volume) such that each tree genotype was separately grown in potting mix with an inoculum of its own genotype's soil ("home" soil) and a soil from an individual from each other population ("away" soil; 12 total genotypes - 3 genotypes X 4 populations). A sub-sample of each soil was stored frozen (-80° C) for amplicon sequencing of the microbial community.

Metabolite Sampling

To address the hypothesis that plant genotype will interact with soil microbial communities from across populations to alter root and associated soil rhizosphere metabolomes, samples of fine roots and rhizosphere soil were taken randomly from multiple locations in each pot after growing for ~ nine months in the greenhouse. Roots less than 1 mm in diameter, from 5 to 10 cm below soil surface, were brushed clean of soil with a fine brush, before being rinsed with deionized water (Mueller et al*., in review*). The soil brushed from the roots was saved as the rhizosphere sample. Aboveground biomass was also determined on each individual plant at experiments end by weighing oven-dried plant tissue (48 h at 105° C). Both root and rhizosphere soil samples were frozen using liquid nitrogen, then homogenized and ground to a fine powder with a mortar and pestle in a cold room (~4° C). Aliquots (~100 mg) were transferred to individual sample tubes where all subsequent extraction steps were performed. The metabolite extraction protocol for the root and rhizosphere soil was adapted from a previously reported method (Stough et al. 2016). Specifically, 1.3 mL of a 2:2:1 methanol: acetonitrile: water with formic acid at a final concentration of 0.1M extraction solution was added and shaken on a vortex mixer for 20 minutes at 4° C. Centrifugation (17,000 g) was utilized for 5 minutes to pelletize the sample, followed by the transfer of the supernatant to a clean sample tube. The sample had 200 µL of extraction solution added and the sample was subjected to the same extraction conditions previously stated. The two supernatants were combined and were

evaporated using nitrogen gas. Sterile water at 4° C was used to re-suspend the dry extract before UPLC-HRMS analysis.

Ultra Performance Liquid Chromatography- High Resolution Mass Spectrometry.

Separation of the metabolites was performed on a Dionex Ultimate liquid chromatography system using a Synergi 2.5µ Hydro-RP100Å, 100 mm x 2.00 mm column (Phenomenex, Torrance, CA, USA). A similar elution gradient utilizing acidic water with tributylamine and methanol over 25 minutes was employed (Lu et al. 2010) The eluent was immediately introduced to an Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA) via electrospray ionization (ESI) with a capillary temperature of 300° C, spray voltage of 3 kV, nitrogen sheath and sweep gas at 25 and 3 units respectively. Data acquisition was done in negative ion mode over the range of 72 – 1000 *m/z* at 140,000 resolution with automatic gain control of $3x10^6$ ions.

Raw data files were converted with MSConvertGUI from Proteowizard (Holman et al. 2014) then uploaded to MAVEN (Clasquin et al. 2012). Metabolites were annotated using exact mass of the $[M-H]^-$ ion and known retention times generated from an in-house curated database in the Biological and Small Molecular Mass Spectrometry Core at University of Tennessee (https://chem.utk.edu/facilities/biological-and-small-molecule-mass-spectrometry-core-bsmmsc/). Area under the curves were compiled for all samples for each metabolite simultaneously and data were normalized to the wet mass of soil or roots extracted.

Unknown Mass Spectrometry Data Reduction.

Spectral features (*m/z*-retention-time pair) were identified by XCMS with a ±5 ppm error window (Tautenhahn et al. 2012), and the CAMERA package (Kuhl et al. 2012) in R (R Development Core Team, 2016) was used to identify potential isotopes and adducts. For a feature to be included in the molecular formula analysis, it must have a signal that is three times that of the blank and must be present in all the replicates in at least one of the sample groups. To avoid weighting the data towards compounds that were detected as multiple chemical species, features were removed that were annotated as the $[M+n]^-$ isotope or identified as an adduct. This reduced spectral feature dataset was used as an input to the Seven Golden Rules (Kind and Fiehn 2007) to generate potential molecular formulas. The formulas were restricted to a mass accuracy

of ±5 ppm and to the following elements: carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S).

Statistical Analysis

Due to mortality, the total planted samples were reduced from 120 (4 populations X 3 genotypes/population X 10 individuals) to 58. Further loss of root tissues in the lab reduced that portion of the dataset to 22 samples (Table 3.1). Using the remaining samples, we used a Jaccard distance-based redundancy analysis (RDA) to assign variation in the root matrix to the four tree populations. Sub-setting by root inoculation was only possible in the trees from the Blue River, where a redundancy analysis was used to determine if soil inoculation had an effect on root metabolites. These statistical tests were also done for the soil metabolite data. While the soil samples had better coverage across tree and soil populations, these samples too, lack the power to detect significant individual effects at the tree X soil level (Fitzpatrick 2009). To examine patterns present in the dataset despite these limitations, the metabolite variation was visualized with methods accounting for population variation (RDA) and juxtaposed with null models (PCA). Compounds responsible for determining the variation among groups were identified with indicator species analysis, where significant indicators were assigned to groups (either tree population or soil inoculum depending on the analysis) based on a permutation test with an α level of 0.01 (De Cáceres et al. 2010). These indicators were then compared to the potential molecular formulas given by the Seven Golden Rules output, filtered by compounds in the PubChem database (U.S. National Library of Medicine). These formulas were then visualized with Van Krevelen diagrams to estimate the chemical families of the different indicator compounds, based on the H:C and O:C ratios of compounds in the chemical mixture (Brockman et al. 2018).

All statistical models and figures were made using R (R Development Core Team, 2016). The redundancy analysis, principal components analysis and indicator species analysis were performed with the package vegan (Oksanen et al. 2016).

Results

Supporting the first hypothesis, tree populations produce unique root tissue metabolomes, with tree population explaining 24% of the variation in the 9,202 known and unknown root metabolites identified, regardless of soil population origin in which they are grown (Table 3.1, Figure 3.1). Soil inoculation did not show an overall difference in root chemistry; models incorporating soil inoculation origin did not perform significantly better than null models (Table 3.1, Figure 3.1). An interaction between tree population and soil inoculation origin suggests that once the variation due to tree population is accounted for, the effect of specific soil microbiomes on trees becomes evident (Table 3.1). At the broad scale therefore, tree population is a primary driver of root metabolites. However, at the population scale, different soil microbiomes can alter the tree root's expressed metabolites, at least at one site (Table 3.2, Blue River). Further evidence for this plant-soil connection can be seen in comparing the strong population signal in "home" soils versus the relatively weak signal of tree population in "away" soils (Table 3.1). Plant populations differed in the number of unique compounds found in their root tissue. Indicator analyses showed that of 9,202 root tissue metabolites 546 were important for differentiating tree populations. The Weber River individuals have the fewest unique metabolites (35), while the Snake River population is associated with the most unique root metabolites (275, Table 3.3).

There is no support for the hypothesized relationship among the community of 5,714 rhizosphere soil metabolites with tree population, soil population origin, or the interaction between tree population and soil inoculation (Table 3.1). Even when restricting the measure of soil metabolites to those individuals growing in "home" soil, there is no pattern of tree population on rhizosphere soil metabolites (Table 3.1); in contrast to previous work (Mueller et al. *in review*). Visualizing the variation of soil metabolites of each tree population however, shows patterns not evident by the p-values. As above, once the influence of the tree population is separated, the effect of the soil inoculation on the rhizosphere soil metabolites is different from null models visually, but not statistically (Table 3.2, Figure 3.2). Despite the limitations of sample size, these results suggest that the rhizosphere soil metabolome can be better described by the interaction of plant species and soil inoculation than suggested by the full model (Table

3.1). For example, when the rhizosphere soil metabolites of samples that were planted with Oak Creek trees are examined, it is the "home" soil condition (Oak Creek soil inoculation) that is the most distinct group (Figure 3.2c), a relationship that the null model failed to detect (Figure 3.2d). Indicator species analysis was used to determine the proportion of metabolites different among tree populations, and the differences among soil inoculations within each tree population. Table 3.3 shows the breakdown of these indicators by tree population and soil inoculum source. A portion of the unknown soil metabolites were identified with the 7 Golden Rules program and graphed with a Van Krevelen diagram to estimate their chemical families (Figure 3.2). Interestingly, the soil samples had an order of magnitude fewer indicator metabolites than the root tissues. Furthermore, the Oak Creek and Weber River soil inoculations had consistently higher unique indicators compared to the Blue River, in contrast to the pattern seen in the root tissues. When Blue River soil inocula create significant indicator compounds (in the cases of Blue River and Snake River tree population subsets), they behave very differently. In Blue River trees, Blue River soils are more saturated (comprised of a higher H/C ratio) than those same soil inocula in Snake River trees (Figure 3.2 c, l).

Discussion

These results support past work that plant intra-specific, population-level, variation predicts the structure of root tissue metabolite communities. We find support for Hypothesis 1, with almost 60% of the variation in root metabolites predicted by the interaction between tree population and soil inoculation. Despite the low sample size inhibiting full exploration of this pattern, indicator species analysis revealed the drastic change in the number of important indicators from the lowest in the Weber River (35) to almost eight times that in the Snake River roots (275). Our analysis does not support Hypothesis 2; the rhizosphere soil metabolome is more variable than can be predicted by tree population or soil inoculation. However, when examined visually, the differing effect of soil inoculation origin on trees from different populations suggests that the relationship between plant population and soil environment exists and it is the lack of statistical power in the interaction model, not the effect size that determines our failure to reject the null hypothesis (Fitzpatrick 2009). This claim is further supported by the variation in indicator compounds by both tree population and soil inoculum. In contrast to previous rhizosphere metabolomics work in *P. angustifolia*, the root metabolite richness was

33% lower than past studies, growing in similar conditions, albeit without a live soil community inoculum (Mueller et al. *in review*). Second, this study showed weaker support for linkage between the surrounding rhizosphere soil metabolite matrix and the plant root. A single year of growth, is likely a key factor explaining the differences of this study relative to our previous work as root age has been shown to alter the metabolites exuded (Chaparro et al. 2013) and plant conditioning of soil communities strengthens over time (Lau and Lennon 2012). Despite this, our data suggest that tree population is the best single predictor of root tissue metabolites, and that the interaction effect of soil inocula on the rhizosphere differ based on plant population.

This work and others have shown the importance of adaptation to the intra-specific trait variation found in *Populus angustifolia* across populations (Smith et al. 2012, Pregitzer et al. 2013, Evans et al. 2015, Ware et al. 2019b, Mueller et al.,*in review*). Genetic variation among populations is in part due to the selective pressures of climatic variation from AZ to MT on plant metabolism and function (Van Nuland et al. 2018, Ware et al. 2019b) as well as the role that variation in soils can have on tree genotypes (Smith et al., 2012). For example, the 10.4° C MAT gradient present across the range selects for variation in bud break phenology, with a 28 day difference in bud break seen in a homogeneous greenhouse environment (Ware et al. 2019b). These differences in aboveground traits should have belowground ecosystem consequences that are seen in many systems (Lojewski et al. 2012, Pregitzer et al. 2013, Mueller et al. 2017). Specifically in plant metabolites, genetically based variation in secondary metabolites was found in Chinese cork oak (*Quercus variabilis*) to vary with latitude changes which encompassed different climates and rates of herbivory (Wang et al., 2016), demonstrating the evolution of plant metabolites along environmental gradients. Other metabolites have been shown to change across abiotic gradients, for example, seasonal variation in temperature and precipitation were related to changes in leaf and stem secondary metabolites (Sampaio et al. 2016). Similarly, plant flavonoids (Winkel-Shirley 2002), and saponins (Szakiel et al. 2011) respond to abiotic and biotic stress, all of which show the response of specific plant compounds to environmental gradients are well documented. Our ability to examine the entire chemical metabolome, including unidentified compounds, is opening up new doors at the interface of ecosystem ecology and analytical chemistry.

Plant root exudates have been linked to signaling of beneficial microbes (Kiers and Denison 2008), direct defense (Baetz and Martinoia 2014), signaling between plants (Chamberlain et al. 2001) as well as allelochemicals reducing competition (Bertin et al. 2003, Hu et al. 2018). Root exudation is a primary way plants shape their surroundings, as it is an important pathway of carbon entering belowground food webs (Pausch and Kuzyakov 2018). In this study, plant population alone structured the variation in root tissue metabolites but in rhizosphere soil metabolites the interaction with soil inoculation was needed to see patterns. However, a number of indicator metabolites found in the rhizosphere soil samples doubled from 11 in pots with Blue River trees to 20 in pots with Snake River trees suggesting that tree root exudate differences between the populations are driving the chemical complexity expressed in the rhizosphere. Visualizations with Van Krevelen diagrams reveals patterns not evident from our statistics. First, not all identical soil inoculations behave similarly when introduced to different plant species. In the extreme case, Blue River soil inoculum was only an important driver of soil metabolite variation when paired with two of the four tree populations, and differed in those two groups. Under Blue River trees, the Blue River soil inoculation identifiers was more saturated (had a higher H/C ratio) than under Snake River trees. Interestingly, this pattern remains for the other inoculations as well. The unique Snake River soil inoculation metabolites are more saturated under their home trees and the nearby Weber River trees than the indicators of that same inoculation under the southern and further away Blue River and Oak Creek trees. These results suggest that the carbon input from trees regulates the effect of soil microbes, which is also supported by the variation in Figure 3.2, showing the important tree population soil X soil inoculation effects. If all plants alter surrounding rhizosphere soils differently based on soil conditions (as seen here based on home/away soil inoculations) understanding the chemical pathways that govern the plant-soil metabolite interaction and the ecosystem consequences of that interaction is the next step for modeling and investigating plant distribution and migration. This then highlights again the importance of generating metabolomics libraries which can better identify the unknown metabolites in these samples. If we want to understand which cellular pathways are changing when plants colonize novel soils, we need to first identify exactly which compounds to examine. Luckily, the methods and results presented here give a good starting point for that search. Examining the differences in indicators using Van Krevelen diagrams suggests that differences in the H/C ratio in identical rhizosphere soil starting conditions are

associated with genetic variations in plant root chemistry, which may have important ecosystem effects.

Conclusion

While reciprocal transplants are the best way to test how migrating tree populations will alter and condition their new soils (Van Nuland et al. 2017), this study shows that tree genetics are an important part of understanding how belowground interactions alter the chemical signals of all involved. Here, plant populations maintained distinct root metabolomes despite soil inoculations from locations more than 1,300 km away from their home sites. Differences in the soil metabolites were also best understood when visualized through the filter of tree population. Despite the small sample size, these results suggest that intraspecific variation in the carbon source alters the entire rhizosphere metabolome. As tree populations move into novel soils in response to global change, the degree to which plants can alter unique soil conditions to benefit themselves will be an important trait to examine. Similar to that found in Mueller et al. *in review,* the unidentified metabolites are necessary for finding meaningful patterns in this dataset. All of the identifying compounds for each tree population and soil inoculation were unidentified from the libraries currently used. While it is clear that focus should be put into building chemical libraries that are better suited for environmental metabolomics, this study shows that in the meantime, we need to utilize the majority of our metabolomics datasets, including the unknowns.

The realization that within-species variation accounts for a large proportion of the biodiversity that we often mis-associate to the species level (Read et al. 2016), and that this trait variation interacts in unique ways with the environment belowground (Pregitzer et al. 2013) means that the ecosystem functions we examine on the landscape are the result of localized interactions between populations of trees and soil communities. Therefore the modern interpretations of the gene by ecosystem framework are necessary to predict the distribution of ecosystem processes (Van Nuland et al. 2016, Ware et al. 2019a). Understanding the chemical mechanisms that drive these gene X ecosystem relationships are a necessary step in helping model the distribution of traits across a landscape. Increasing evidence of intraspecific root metabolite variation (Hu et al. 2018, Mueller et al., *in review*) makes it clear that populations can evolve distinct rhizosphere metabolite communities, and our work here suggests that these chemical communities are driven primarily by autotrophic inputs controlled by plant genetics.

Accounting for this variation and its interaction with soil microbial communities will be key to predicting and modelling the distribution of ecosystem function by showing the importance of population, phenotype, and resource by phenotype approaches in a changing world (Marcer et al. 2016, Kivlin et al. 2017, Ikeda et al. 2017, Benito Garzón et al. 2019).

Acknowledgements

Thanks to Philip Patterson at the Northern Arizona University greenhouses and to Ken McFarland and Jeff Martin for their greenhouse expertise at the University of Tennessee. Special thank you to Melissa Liotta for assistance with figures. We acknowledge funding from The University of Tennessee, Department of Ecology and Evolutionary Biology.

References

Badri, D. V., and J. M. Vivanco. 2009. Regulation and function of root exudates. Plant, Cell & Environment 32:666–681.

Badri, D. V., T. L. Weir, D. van der Lelie, and J. M. Vivanco. 2009. Rhizosphere chemical dialogues: plant–microbe interactions. Current Opinion in Biotechnology 20:642–650.

Baetz, U., and E. Martinoia. 2014. Root exudates: the hidden part of plant defense. Trends in Plant Science 19:90–98.

Bais, H. P., T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Annual Review of Plant Biology 57:233–266.

Bardgett, R. D., and D. A. Wardle. 2010. Aboveground-Belowground Linkages Biotic Interactions, Ecosystem Processes, and Global Change. Oxford University Press, Oxford.

Benito Garzón, M., T. M. Robson, and A. Hampe. 2019. ΔTrait <scp>SDMs</scp> : species distribution models that account for local adaptation and phenotypic plasticity. New Phytologist 222:1757–1765.

Bertin, C., X. Yang, and L. A. Weston. 2003. The role of root exudates and allelochemicals in the rhizosphere. Plant and Soil 256:67–83.

Bezemer, T. M., and N. M. van Dam. 2005. Linking aboveground and belowground interactions

via induced plant defenses. Trends in Ecology & Evolution 20:617–624.

Bezemer, T. M., C. S. Lawson, K. Hedlund, A. R. Edwards, A. J. Brook, J. M. Igual, S. R. Mortimer, and W. H. Van Der Putten. 2006. Plant species and functional group effects on abiotic and microbial soil properties and plant-soil feedback responses in two grasslands. Journal of Ecology 94:893–904.

Brady, K. U., A. R. Kruckeberg, and H. D. Bradshaw Jr. 2005. Evolutionary Ecology of Plant Adaptation to Serpentine Soils. Annual Review of Ecology, Evolution, and Systematics 36:243–266.

Chagas, F. O., R. D. C. Pessotti, A. M. Caraballo-Rodríguez, and M. T. Pupo. 2018. Chemical signaling involved in plant–microbe interactions. Chemical Society Reviews 47:1652–1704.

Chamberlain, K., E. Guerrieri, F. Pennacchio, J. Pettersson, J. A. Pickett, G. M. Poppy, W. Powell, L. J. Wadhams, and C. M. Woodcock. 2001. Can aphid-induced plant signals be transmitted aerially and through the rhizosphere? Biochemical Systematics and Ecology 29:1063–1074.

Chaparro, J. M., D. V. Badri, M. G. Bakker, A. Sugiyama, D. K. Manter, and J. M. Vivanco. 2013. Root Exudation of Phytochemicals in Arabidopsis Follows Specific Patterns That Are Developmentally Programmed and Correlate with Soil Microbial Functions. PLoS ONE 8:e55731.

Chapin, F. S., K. Autumn, F. Pugnaire, and F. S. Chapin. 1993. Evolution of Suites of Traits in Response to Environmental Stress characteristic. The American Naturalist 142:78–92.

Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Current protocols in bioinformatics Chapter 14:Unit14.11.

van Dam, N. M., and H. J. Bouwmeester. 2016. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. Trends in Plant Science 21:256–265.

Dennis, P. G., A. J. Miller, and P. R. Hirsch. 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? FEMS Microbiology Ecology 72:313–327.

Evans, L. M., G. J. Allan, S. P. Difazio, G. T. Slavov, J. A. Wilder, K. D. Floate, S. B. Rood, and T. G. Whitham. 2015. Geographical barriers and climate influence demographic history in narrowleaf cottonwoods. Heredity 114:387–396.

Fitzpatrick, B. M. 2009. Power and sample size for nested analysis of molecular variance. Molecular Ecology 18:3961–3966.

Grayston, S. J., D. Vaughan, and D. Jones. 1997. Rhizosphere carbon flow in trees, in comparison with annual plants: the importance of root exudation and its impact on microbial activity and nutrient availability. Applied Soil Ecology 5:29–56.

Harrison, S., H. D. Safford, J. B. Grace, J. H. Viers, and K. F. Davies. 2006. Regional and local species richness in an insular environment: Serpentine plants in California. Ecological Monographs 76:41–56.

Hendry, A. P. 2019. A critique for eco-evolutionary dynamics. Functional Ecology 33:84–94.

Holman, J. D., D. L. Tabb, and P. Mallick. 2014. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. Pages 13.24.1-13.24.9 Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Hu, L., C. A. M. Robert, S. Cadot, X. Zhang, M. Ye, B. Li, D. Manzo, N. Chervet, T. Steinger, M. G. A. van der Heijden, K. Schlaeppi, and M. Erb. 2018. Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. Nature Communications 9:2738.

Ikeda, D. H., T. L. Max, G. J. Allan, M. K. Lau, S. M. Shuster, and T. G. Whitham. 2017. Genetically informed ecological niche models improve climate change predictions. Global Change Biology 23:164–176.

Kardol, P., T. Martijn Bezemer, and W. H. Van Der Putten. 2006. Temporal variation in plant-soil feedback controls succession. Ecology Letters 9:1080–1088.

Kiers, E. T., and R. F. Denison. 2008. Sanctions, Cooperation, and the Stability of Plant-Rhizosphere Mutualisms. Annual Review of Ecology, Evolution, and Systematics 39:215–236.

Kind, T., and O. Fiehn. 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8:105.

Kivlin, S. N., R. Muscarella, C. V. Hawkes, and K. K. Treseder. 2017. The Predictive Power of Ecological Niche Modeling for Global Arbuscular Mycorrhizal Fungal Biogeography. Pages 143–158.

Kudoyarova, G. R., I. C. Dodd, D. S. Veselov, S. A. Rothwell, and S. Yu. Veselov. 2015. Common and specific responses to availability of mineral nutrients and water. Journal of Experimental Botany 66:2133–2144.

Kuhl, C., R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. 2012. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry 84:283–289.

Lau, J. A., and J. T. Lennon. 2012. Rapid responses of soil microorganisms improve plant fitness in novel environments. Proceedings of the National Academy of Sciences 109:14058–14062.

Lojewski, N. R., D. G. Fischer, J. K. Bailey, J. A. Schweitzer, T. G. Whitham, and S. C. Hart. 2012. Genetic components to belowground carbon fluxes in a riparian forest ecosystem: A common garden approach. New Phytologist 195:631–639.

Lu, W., M. F. Clasquin, E. Melamud, D. Amador-Noguez, A. A. Caudy, and J. D. Rabinowitz. 2010. Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone orbitrap mass spectrometer. Analytical Chemistry 82:3212–3221.

Lugtenberg, B. J. J., L. V. Kravchenko, and M. Simons. 1999. Tomato seed and root exudate sugars: composition, utilization by Pseudomonas biocontrol strains and role in rhizosphere colonization. Environmental Microbiology 1:439–446.

Mao, Y., X. Li, E. M. Smyth, A. C. Yannarell, and R. I. Mackie. 2014. Enrichment of specific bacterial and eukaryotic microbes in the rhizosphere of switchgrass (P anicum virgatum L.) through root exudates. Environmental Microbiology Reports 6:293–306.

Marcer, A., B. Méndez-Vigo, C. Alonso-Blanco, and F. X. Picó. 2016. Tackling intraspecific genetic structure in distribution models better reflects species geographical range. Ecology

and Evolution 6:2084–2097.

Marti, G., M. Erb, J. Boccard, G. Glauser, G. R. Doyen, N. Villard, C. A. M. Robert, T. C. J. Turlings, S. Rudaz, and J. L. Wolfender. 2013. Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. Plant, Cell and Environment 36:621–639.

Martijn Bezemer, T., W. H. van der Putten, H. Martens, T. F. J. van de Voorde, P. P. J. Mulder, and O. Kostenko. 2013. Above- and below-ground herbivory effects on below-ground plant-fungus interactions and plant-soil feedback responses. Journal of Ecology 101:325–333.

Metcalfe, D. B., R. A. Fisher, and D. A. Wardle. 2011. Plant communities as drivers of soil respiration: pathways, mechanisms, and significance for global change. Biogeosciences 8:2047–2061.

Mhlongo, M. I., L. A. Piater, N. E. Madala, N. Labuschagne, and I. A. Dubery. 2018. The Chemistry of Plant–Microbe Interactions in the Rhizosphere and the Potential for Metabolomics to Reveal Signaling Related to Defense Priming and Induced Systemic Resistance. Frontiers in Plant Science 9:112.

Michalet, S., J. Rohr, D. Warshan, C. Bardon, J. C. Roggy, A. M. Domenach, S. Czarnes, T. Pommier, B. Combourieu, N. Guillaumaud, F. Bellvert, G. Comte, and F. Poly. 2013. Phytochemical analysis of mature tree root exudates in situ and their role in shaping soil microbial communities in relation to tree N-acquisition strategy. Plant Physiology and Biochemistry 72:169–177.

Mueller, L. O., S. R. Borstein, E. D. Tague, S. P. Dearth, H. F. Castro, S. R. Campagna, J. K. Bailey, and J. A. Schweitzer. (*in review*). Populations of Populus angustifolia have evolved distinct metabolic profiles that influence their surrounding soil. Plant and Soil *in revirew.*

Mueller, L. O., L. C. Breza, M. A. Genung, C. P. Giardina, N. E. Stone, L. C. Sidak-Loftis, J. D. Busch, D. M. Wagner, J. K. Bailey, and J. A. Schweitzer. 2017. Ecosystem consequences of plant genetic divergence with colonization of new habitat. Ecosphere 8:e01743.

Van Nuland, M. E., J. K. Bailey, and J. A. Schweitzer. 2017. Divergent plant–soil feedbacks

could alter future elevation ranges and ecosystem dynamics. Nature Ecology & Evolution 1:0150.

Van Nuland, M. E., I. M. Ware, J. K. Bailey, and J. A. Schweitzer. 2018. Geographic variation in plant evolutionary responses to soil fertility is mediated by soil microbial communities. Functional Ecology 33:95–106.

Van Nuland, M. E., R. C. Wooliver, A. A. Pfennigwerth, Q. D. Read, I. M. Ware, L. Mueller, J. A. Fordyce, J. A. Schweitzer, and J. K. Bailey. 2016. Plant-soil feedbacks: connecting ecosystem ecology and evolution. Functional Ecology 30:1032–1042.

Oksanen, A. J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. Mcglinn, P. R. Minchin, R. B. O. Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2016. Vegan: Community Ecology Package. https://github.com/vegandevs/vegan.

Orwin, K. H., S. M. Buckland, D. Johnson, B. L. Turner, S. Smart, S. Oakley, and R. D. Bardgett. 2010. Linkages of plant traits to soil properties and the functioning of temperate grassland. Journal of Ecology 98:1074–1083.

Panke-Buisse, K., S. Lee, and J. Kao-Kniffin. 2017. Cultivated Sub-Populations of Soil Microbiomes Retain Early Flowering Plant Trait. Microbial Ecology 73:394–403.

Panke-Buisse, K., A. C. Poole, J. K. Goodrich, R. E. Ley, and J. Kao-Kniffin. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. The ISME Journal 9:980–989.

Pausch, J., and Y. Kuzyakov. 2018. Carbon input by roots into the soil: Quantification of rhizodeposition from root to ecosystem scale. Global Change Biology 24:1–12.

Pétriacq, P., A. Williams, A. Cotton, A. E. McFarlane, S. A. Rolfe, and J. Ton. 2017. Metabolite profiling of non-sterile rhizosphere soil. Plant Journal 92:147–162.

Pregitzer, C. C., J. K. Bailey, S. C. Hart, and J. A. Schweitzer. 2010. Soils as agents of selection: feedbacks between plants and soils alter seedling survival and performance. Evolutionary Ecology 24:1045–1059.

Pregitzer, C. C., J. K. Bailey, and J. A. Schweitzer. 2013. Genetic by environment interactions

affect plant-soil linkages. Ecology and Evolution 3:2322–2333.

Read, Q. D., S. M. Hoban, M. B. Eppinga, J. A. Schweitzer, and J. K. Bailey. 2016. Accounting for the nested nature of genetic variation across levels of organization improves our understanding of biodiversity and community ecology. Oikos 125:895–904.

Rellán-Álvarez, R., G. Lobet, and J. R. Dinneny. 2016. Environmental Control of Root System Biology. Annual Review of Plant Biology 67:619–642.

Ristok, C., Y. Poeschl, J.-H. Dudenhöffer, A. Ebeling, N. Eisenhauer, F. Vergara, C. Wagg, N. M. van Dam, and A. Weinhold. 2019. Plant species richness elicits changes in the metabolome of grassland species via soil biotic legacy. Journal of Ecology Early view.

Rovira, A. D. 1969. Plant root exudates. The Botanical Review 35:35–57.

Sampaio, B. L., R. Edrada-Ebel, and F. B. Da Costa. 2016. Effect of the environment on the secondary metabolic profile of Tithonia diversifolia: a model for environmental metabolomics of plants. Scientific Reports 6:29265.

Smith, D. S., J. A. Schweitzer, P. Turk, J. K. Bailey, S. C. Hart, S. M. Shuster, and T. G. Whitham. 2012. Soil-mediated local adaptation alters seedling survival and performance. Plant and Soil 352:243–251.

Souza, R. de, A. Ambrosini, and L. M. P. Passaglia. 2015. Plant growth-promoting bacteria as inoculants in agricultural soils. Genetics and Molecular Biology 38:401–419.

Stough, J. M. A., S. P. Dearth, J. E. Denny, G. R. LeCleir, N. W. Schmidt, S. R. Campagna, and S. W. Wilhelm. 2016. Functional Characteristics of the Gut Microbiome in C57BL/6 Mice Differentially Susceptible to Plasmodium yoelii. Frontiers in Microbiology 7:1520.

Swenson, T. L., U. Karaoz, J. M. Swenson, B. P. Bowen, and T. R. Northen. 2018. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. Nature Communications 9:19.

Szakiel, A., C. Pączkowski, and M. Henry. 2011. Influence of environmental biotic factors on the content of saponins in plants. Phytochemistry Reviews 10:493–502.

Tautenhahn, R., G. J. Patti, D. Rinehart, and G. Siuzdak. 2012. XCMS online: A web-based

platform to process untargeted metabolomic data. Analytical Chemistry 84:5035–5039.

Team, R. D. C., and R. R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Treseder, K. K., and P. M. Vitousek. 2001. Potential ecosystem-level effects of genetic variation among populations of Metrosideros polymorpha from a soil fertility gradient in Hawaii. Oecologia 126:266–275.

van der Putten, W. H. 2003. Plant defense belowground and spatiotemporal processes in natural vegetation. Ecology 84:2269–2280.

van der Putten, W. H., M. A. Bradford, E. Pernilla Brinkman, T. F. J. van de Voorde, and G. F. Veen. 2016. Where, when and how plant–soil feedback matters in a changing world. Functional Ecology 30:1109–1121.

Vitousek, P. M., L. R. Walker, L. D. Whiteaker, D. Mueller-Dombois, and P. A. Matson. 1987. Biological Invasion by Myrica faya Alters Ecosystem Development in Hawaii. Science 238:802–804.

Ware, I. M., C. R. Fitzpatrick, A. Senthilnathan, S. L. J. Bayliss, K. K. Beals, L. O. Mueller, J. L. Summers, R. C. Wooliver, M. E. Van Nuland, M. T. Kinnison, E. P. Palkovacs, J. A. Schweitzer, and J. K. Bailey. 2019a. Feedbacks link ecosystem ecology and evolution across spatial and temporal scales: Empirical evidence and future directions. Functional Ecology 33:31–42.

Ware, I. M., M. E. Van Nuland, J. A. Schweitzer, Z. Yang, C. W. Schadt, L. C. Sidak-Loftis, N. E. Stone, J. D. Busch, D. M. Wagner, and J. K. Bailey. 2019b. Climate-driven reduction of genetic variation in plant phenology alters soil communities and nutrient pools. Global Change Biology 25:1514–1528.

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. LeRoy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, B. M. Potts, D. G. Fischer, C. A. Gehring, R. L. Lindroth, J. C. Marks, S. C. Hart, G. M. Wimp, and S. C. Wooley. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics 7:510–523.

Winkel-Shirley, B. 2002. Biosynthesis of flavonoids and effects of stress. Current Opinion in Plant Biology 5:218–223.

Wolfe, B. E., V. L. Rodgers, K. A. Stinson, and A. Pringle. 2008. The invasive plant Alliaria petiolata (garlic mustard) inhibits ectomycorrhizal fungi in its introduced range. Journal of Ecology 96:777–783.

Wooliver, R. C., J. K. Senior, B. M. Potts, M. E. Van Nuland, J. K. Bailey, and J. A. Schweitzer. 2018. Soil fungi underlie a phylogenetic pattern in plant growth responses to nitrogen enrichment. Journal of Ecology 106:2161–2175.

Wurst, S., R. Wagenaar, A. Biere, and W. H. van der Putten. 2010. Microorganisms and nematodes increase levels of secondary metabolites in roots and root exudates of Plantago lanceolata. Plant and Soil 329:117–126.

Zhalnina, K., K. B. Louie, Z. Hao, N. Mansoori, U. N. da Rocha, S. Shi, H. Cho, U. Karaoz, D. Loqué, B. P. Bowen, M. K. Firestone, T. R. Northen, and E. L. Brodie. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nature Microbiology 3:470–480.

Appendix

Table 3.1: Differences in root tissue and rhizosphere soil sample metabolites explored with redundancy analysis (RDA). Proportion explained represents all RDA axes (3 for single effect models, 11 for interaction models). P-values were generated by permutation when compared to null models. Bolded p-values are where $p \leq \alpha$ (α=0.05). The sampling subset represents which rows of the data matrix were included. The "Constrained by" column identifies which groups the RDA used as fixed effects.

Sampling sub-set	Constrained by:	Sample size	Proportion explained	P value
		Root Tissue Metabolites		
All root tissue	tree population	22	0.2426	**0.003**
All root tissue	soil inoculation	22	0.1413	0.43
All root tissue	Tree population * soil inoculation	22	0.5937	**0.038**
"home" soil	Tree population	5	0.7595	0.0667
"away" soil	Tree population	17	0.1696	0.741
		Rhizosphere Soil Metabolites		
All Rhizosphere soils	Tree population	58	0.1054	0.453
All Rhizosphere soils	Soil inoculation	58	0.1272	0.676
All Rhizosphere soils	Tree population * soil inoculation	58	0.1162	0.144
"home" soil	Tree population	15	0.2231	0.611
"away" soil	Tree population	43	0.0698	0.477

Table 3.2: Differences in root tissue and rhizosphere soil sample metabolites caused by the soil inocula origin for each tree population, explored with redundancy analysis (RDA). Proportion explained represents all three RDA axes. P-values were generated by permutation when compared to null models. The sampling subset represents which tree population is included for that model. All models were constrained by the fixed effect of soil inoculation treatment. Sample size is a function of both remaining samples and the subset of the data needed for an analysis. The mathematical reasoning behind the lack of significant p-values despite high predictive intervals of the models is described in Fitzpatrick, (2009).

Sampling subset	Constrained by:	Sample size	Proportion explained	P value
Root Tissue Metabolites				
Blue River	soil inoculation	9	0.4542	0.076
Oak Creek	soil inoculation	4	0.5077	0.999
Weber River	soil inoculation	6	0.4315	0.400
Snake River	soil inoculation	3	0.4858	0.667
Rhizosphere Soil Metabolites				
Blue River	soil inoculation	16	0.1617	0.919
Oak Creek	soil inoculation	13	0.1774	0.894
Weber River	soil inoculation	15	0.1977	0.712
Snake River	soil inoculation	11	0.2407	0.880

Table 3.3: Numbers of unique compounds associated with the differences in tree population and soil inoculation origin found via indicator species analysis at α = 0.01. Values represent the number (proportion of total) of metabolites that associate with that tree population or soil inoculation.

Number of indicator compounds separating root tissue metabolites by population					
Data subset	**BL**	**OC**	**WR**	**SNR**	**Total**
All root tissue	161 (29%)	75 (14%)	35 (6%)	275 (50%)	546
Number of indicator compounds separating rhizosphere soil metabolites by root population and soil inoculation source					
Data subset	**BL soil inoculum**	**OC soil inoculum**	**WR soil inoculum**	**SNR soil inoculum**	**Total**
BL Trees	4 (36%)	4 (36%)	2 (18%)	1 (9%)	11
OC Trees	0 (0%)	3 (21%)	3 (21%)	8 (57%)	14
WR Trees	0 (0%)	4 (22%)	10 (56%)	4 (22%)	18
SNR Trees	10 (50%)	6 (30%)	2 (10%)	2 (10%)	20

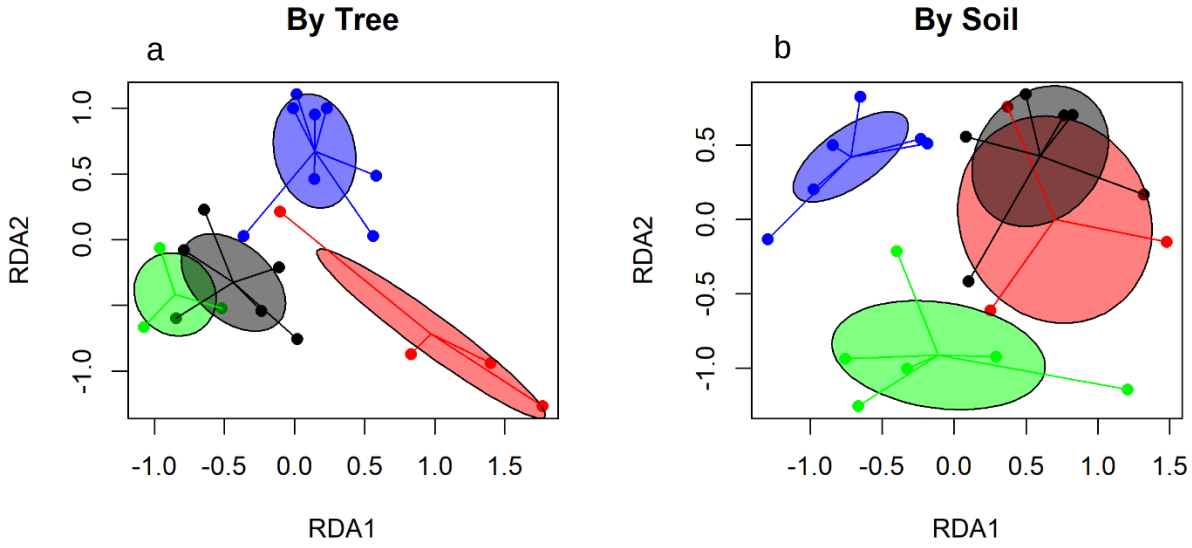
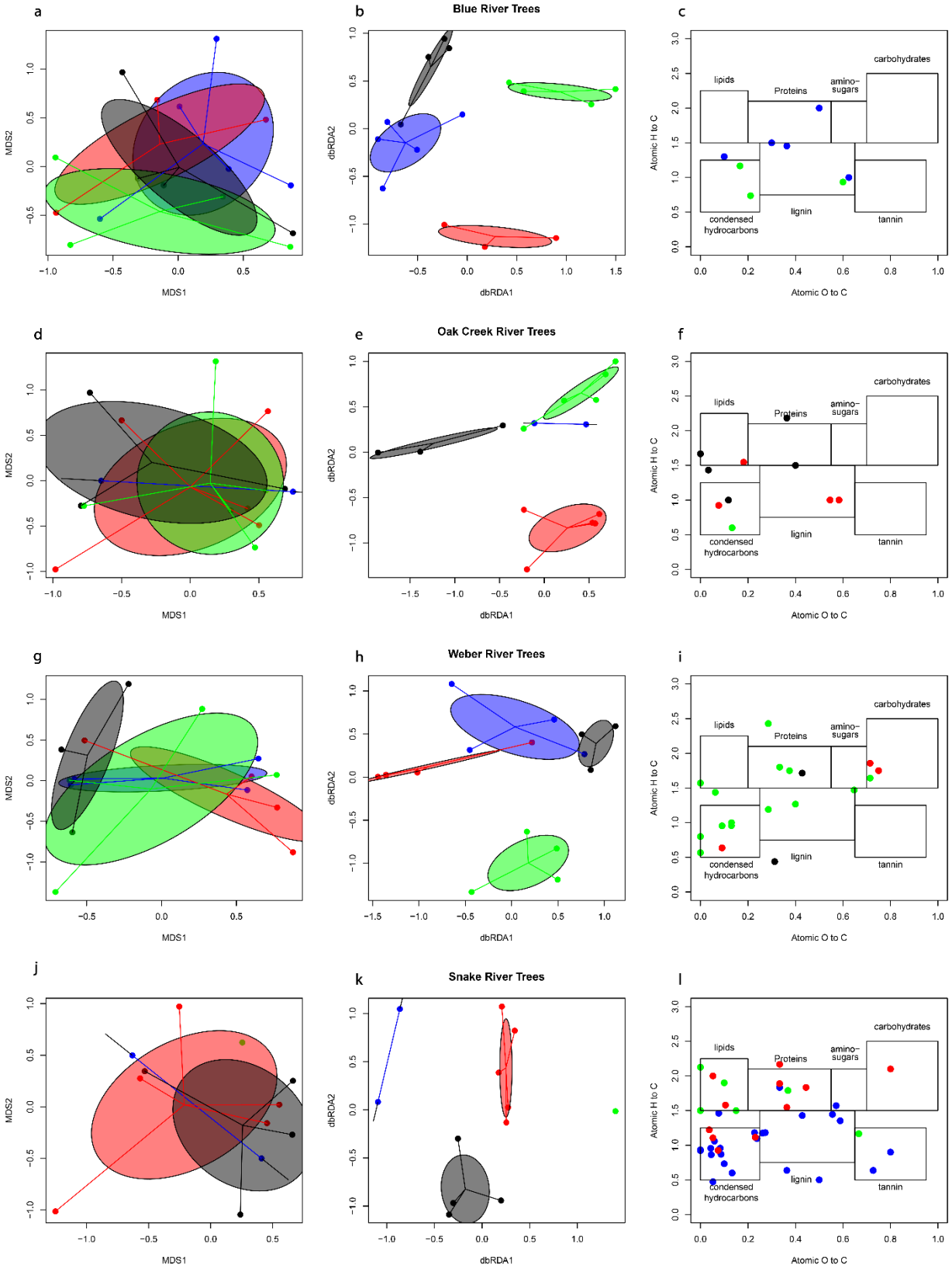


Figure 3.1: Visualization of the separation of root tissue metabolites: by tree population across all soil inocula (a) and soil inocula across all tree populations (b) on the first two constrained RDA axes. Color of the points identifies the tree population of origin (a), and the inoculum collection location; the Snake River (SNR) is represented in green, the Weber River (WR) in black, the Oak Creek (OC) in red, and the Blue River (BL) in blue. Points represent the position of a metabolite sample in multidimensional space, and ellipses represent standard error around each centroid. Tree population was a significant predictor of the variation in root metabolites, while soil inoculation origin was not (See Table 3.1 lines 1 & 2).

Figure 3.2: Separating the soil metabolite data by tree population of origin. Visualization of the relationship between soil metabolites of different soil inoculation origins. Redundancy analysis visualizing the effect of soil inoculation for each tree population are in the middle (b, e, h, k), while null models ignoring the effect of soil inoculation are on the left (a, d, g, j). Van Krevelen diagrams representing an estimate of the O/C and H/C ratios of important metabolites separating the effects of the soil inocula (c, f, i, l). Color of the points identifies the inoculum collection location; the Snake River (SNR) is represented in green, the Weber River (WR) in black, Oak Creek (OC) in red, and the Blue River (BL) in blue. Visual differences between redundancy models and null models can be discerned despite the lack of statistical significance (Table 3.2).


a
b
Blue River Trees
c
d
e
Oak Creek River Trees
f
g
h
Weber River Trees
i
j
k
Snake River Trees
l
MDS1
MDS2
dbRDA1
dbRDA2
Atomic O to C
Atomic H to C
lipids
Proteins
amino-
sugars
carbohydrates
condensed
hydrocarbons
lignin
tannin

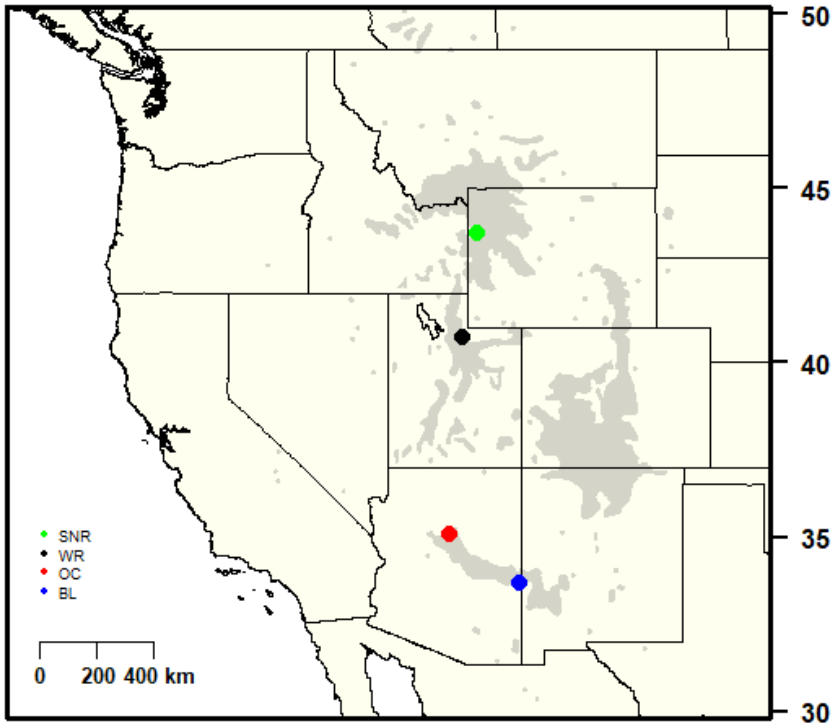


Figure S3.1: Map of the four sampled populations across the distribution of the *P. angustifolia* range (grey outline). The color of the points identifies the collection location; the Snake River (SNR) in green, the Weber River (WR) in black, the Oak Creek (OC) in red, and the Blue River (BL) in blue.

CHAPTER IV

ANALYSIS OF SOIL METABOLOMICS DATA USING MULTI-VARIATE APPROACHES: BEST PRACTICES FOR ECOLOGISTS

Abstract

It is becoming easier and easier to obtain very large data sets associated with soil samples, from metabolomics, genomics, metagenomics, meta-transcriptomics as well as abiotic factors of sites from climate and geologic databases. Despite the growth in data acquisition techniques, statistical methods for analyses of these datasets are often overly simplified. The industry standards for statistical metabolomics analysis and visualization may be applied inconsistently or inappropriately for the more complex questions that metabolomics data sets are being used to address. Furthermore, almost all current metabolomics studies ignore the complexity of the unidentified metabolites in their datasets. Ecological experimental design often has multiple factors, nested levels of hierarchy and large numbers of unique groups. This review examines when a variety of visualization techniques such as non-metric multidimensional scaling, principal components analysis, redundancy analysis, and Van Krevelen diagrams are useful tools. Moreover, we explore how the differences in hypothesis testing techniques including partial least squares-discriminant analysis, redundancy analysis, Mantel tests, and co-inertia alter the interpretations of a single case study metabolite data set. Our goal is to highlight the variety of available statistical tools for metabolomics analysis and describe when each is appropriate to use and encourage methods that can examine unknown compounds.

Introduction

Ecologists have begun to fully embrace metabolomics (the study of secondary cellular compounds of small molecular weight) for analyzing the soil rhizosphere to better understand the chemical linkages of plants, microbes and ecosystem processes. For example, metabolomics has been used recently to examine the effect of plant root exudates on bulk soils (Pétriacq et al. 2017), changes in root exudation based on nutrient accuisition strategy (Michalet et al. 2013), and plant-soil feedbacks (Hu et al. 2018). The reason these approaches are increasingly used in soil ecology is their potential to elucidate many interactions among plants, microbes and soils to better under plant-microbe signals and soil-mediated ecosystem processes. Moreover, specific extraction protocols for soil organic matter are being produced (Swenson and Northen 2019), reviews are being published (van Dam and Bouwmeester 2016, Mhlongo et al. 2018), and metabolomics are beginning to be used in more complex soil matrices (Hu et al., 2018; Mueller et al.*in review*). This is worthwhile, as metabolomics allow for a less biased sampling approach and an enormous increase in chemical data of use to a wide variety of questions, compared with methods focused solely on single compounds. These benefits come with increased chemical complexity that can create analytic challenges. Soil metabolomics datasets provide different sets of challenges than plant tissue metabolomics as these datasets can be large with thousands of unknown compounds that are often omitted and not utilized. Despite the quick adoption of metabolomics analytical techniques in soil ecology, researchers have not adopted a broad set of statistical tools appropriate for making the most out of these data, particularly unidentified metabolites that may be important drivers of ecosystem processes.

The complex datasets of many metabolomics studies are artificially simplified by the statistical approach taken. Most soil metabolomics studies to date have simplified their research question in an attempt to simplify the complex data that metabolomics produce, often using univariate statistical techniques, or only ordination, when the research question would be better suited to other, multivariate statistical approaches. For all but the studies examining the role of a single metabolite, a univariate approach does not fully utilize the power of these datasets or allow for robust analyses with high inference. One of the strengths of metabolomics is the ability to quantify a large chemical community in the sample. Reducing that complexity down to a single compound of interest with a univariate statistical approach negates the reason to explore the metabolome in the first place. Of the published studies examining metabolomes on or related

to soil to date, few have used redundancy analysis or similar multivariate hypothesis testing approaches (but see Hao et al., 2018; Peters et al., 2018; Ristok et al., 2019; Mueller et al.*in review*). Others that have used ordination techniques to visualize metabolites fail to test hypotheses in multivariate space (Wang et al. 2019). Multivariate approaches are a comprehensive means of utilizing all metabolites in a dataset and better capture interactions in the dataset while explicitly testing statistical hypotheses. Often, soil metabolomics are used to answer the following types of questions: 1) What is the difference among the samples? 2) Does an external pressure (i.e. climate, location, lab condition) explain the variation among the samples? 3) Does the chemical complexity of one part of a system influence another (e.g., interactions between plant and soil metabolomes)?

Past reviews of statistical methods involved in metabolomics (Broadhurst and Kell 2007, Liland 2011, Eliasson et al. 2011, Gromski et al. 2015) are written primarily for chemists and do not contain a comprehensive list of appropriate statistical tests and visualizations for the questions outlined above. Many metabolomics papers, particularly in the medical field, have recommended Partial Least Squares Discriminant Analysis (PLS-DA) to examine trends among groups (Eliasson et al. 2011), although with some reservation (Gromski et al. 2015). This test, and other discriminant analyses, are useful tools for identifying individual metabolites that drive differences between groups, but are not distinctly testing the importance of those groups versus others. However, the specific challenges of soil metabolomics, including the relative importance of unknown compounds and the complexity of research questions often require different statistical methods than those common in medical fields. Ordination and visualization techniques such as Non-Metric Multi-Dimensional Scaling (NMDS, Kruskal, 1964), Principal Components Analysis (PCA, Pearson, K., 1901; Hotelling, 1933), Van Krevelen diagrams (Van Krevelen, 1950, which are unique to metabolomics), can all be used to visualize data (See Table 1 for descriptions, limitations and uses of each test). To examine if external independent factors influence a metabolome, ordination and hypothesis testing with distance based Redundancy Analysis (RDA; McArdle & Anderson, 2001) should be considered, and is currently under-utilized in environmental chemistry. If samples can be paired, such as separated horizons of a soil sample, or a root vs. the surrounding rhizosphere, techniques like the Mantel Test (Mantel 1967) and co-inertia (Dolédec & Chessel, 1994) can be useful; the latter being most appropriate with hierarchical datasets (i.e. soil horizons across different canopy covers). These methods

cover the vast majority of cases of interest to soil ecologists to examine the structure of metabolomics samples and address hypotheses based on factors governing that structure. Outside of the scope of this manuscript, more complex techniques (for example, see Principal Coordinates of Neighborhood Matrix, PCNM in Dray & Dufour, 2007, or multivariate random forests (Cutler et al. 2007, Segal and Xiao 2011) can be considered for addressing comparisons with more than two matrices, or better ways of de-tangling important metabolites in multiple groups.

Here, we take a single soil metabolomics dataset (Mueller et al.*in review*) as a case study and compare the assumptions, approaches and outcomes of multiple data visualization tools and multiple multivariate hypothesis-testing techniques (including PLS-DA) to illustrate what results and conclusions can be derived from each to allow researchers to make robust choices to analyze specific research questions. The data from Mueller et al. (*in review*) allow all of these comparisons as they are based on paired root and soil samples from multiple populations of the same species. This dataset also contains both known and unknown metabolites that can be shown as both raw data and distance matrices; these data can be compared as either correlational comparisons or with directional hypothesis testing. This example dataset is not meant to be a definitive test of these statistical methods, but instead should allow a better understanding of the application of different analyses with specific questions as well as appropriate interpretation of results. We also show how and why to incorporate unknown metabolites into analyses to fully explore metabolomics datasets.

Materials and methods

Data acquisition

Data for this comparison were derived from Mueller et al. (*in review*) which examined the role of plant population on soil metabolomic profiles in the root and rhizosphere. Briefly, plant cuttings from six populations of narrowleaf cottonwood (*Populus angustifolia*) from across its range in the intermountain west were collected and grown for four years in a common greenhouse environment. At the end of this period, two samples of fine roots and rhizosphere soil were taken from four genotypes of each of the six populations. Roots from 5 to 10 cm below the soil surface were brushed with a fine brush to remove soil, before being rinsed with deionized water. The soil brushed from the roots was saved as the rhizosphere sample. Both root

and rhizosphere soil samples were frozen using liquid nitrogen, then homogenized and ground to a fine powder with a mortar and pestle in a cold room (~4°C). The metabolomics extraction protocol for the root and rhizosphere soil was adapted from a previously reported method (Stough et al. 2016). Separation of the metabolites was performed on a Dionex Ultimate liquid chromatography system using a Synergi 2.5µ Hydro-RP100Å, 100 mm x 2.00 mm column (Phenomenex, Torrance, CA, USA). Data acquisition was performed in negative ion mode over the range of 72 – 1000 *m/z* at 140,000 resolution with automatic gain control of $3x10^6$ ions. Raw data files were converted with MSConvertGUI from Proteowizard (Holman et al. 2014) then uploaded to MAVEN (Clasquin et al. 2012). Metabolites were annotated using exact mass of the $[M-H]^-$ ion and known retention times generated from an in-house curated database in the Biological and Small Molecular Mass Spectrometry Core at University of Tennessee. Area under the curves were compiled for all samples for each metabolite simultaneously and data was normalized to the wet mass of root or soil extracted.

Spectral features (*m/z*-retention-time pair) were identified by XCMS with a ±5 ppm error window (Tautenhahn et al. 2012), and the CAMERA package (Kuhl et al. 2012) in R (R Development Core Team, 2016) was used to identify potential isotopes and adducts. For a feature to be included in the molecular formula analysis, it must have a signal that is three times that of the blank and must be present in all the replicates in at least one of the sample groups. To avoid weighting the data towards compounds that were detected as multiple chemical species, features were removed that were annotated as the $[M+n]^-$ isotope or identified as an adduct. To explore the ~98% of the dataset that could not be classified, the reduced spectral feature dataset was used as an input to the Seven Golden Rules (Kind and Fiehn 2007) to generate potential molecular formulas. The formulas were restricted to a mass accuracy of ± 5 ppm and to the following elements: carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S).

Data analysis approaches

To explore a range of visualization and analysis techniques, both root and soil samples were used. Like many environmental metabolomics assays with Liquid Chromatography-Mass Spectrometry (LC-MS), the majority of detected metabolites were identified using available libraries. Instead of ignoring the unidentified compounds (98% of the data), all analyses were

undertaken with all available metabolite data. To more easily compare the effects of each test, we chose to display and discuss the root metabolite data in all sections where only one data matrix is being analyzed. The root dataset is 1.5 times larger than the soils and showed stronger patterns among populations, which help demonstrate the potential differences between the analyses. The soil data is used when the visualization or test are appropriate for two metabolomics matrices.

Ordination and data visualization

Ordination methods are techniques used to visualize multi-dimensional data in fewer dimensions, although they generally do not allow for explicit hypothesis testing. Ordination techniques are common in soil metabolomics studies to date, but these methods are not often followed up with multivariate hypothesis testing (Michalet et al. 2013, Wang et al. 2019). Ordination attempts to display relationships among samples in three or fewer axes similar to the distances in the full dimension space. Importantly, these methods can graphically show how similar samples are, or highlight relationships between compounds across samples. If the researcher's goal is to examine patterns obscured by data complexity, ordination is a necessary first step.

A common ordination technique in the field of metabolomics has been NMDS (Wakelin et al., 2015). This method permutes a low stress solution for compressing complex data down to three or fewer axes. The benefit to this approach is that it can be used for multiple data types, including the semi-quantitative data common to metabolomics, generated by relative abundances instead of concentrations. Importantly, NMDS sacrifices the Euclidian distance between points in order to compress the data onto fewer axes, resulting in the position between samples to be different depending on the center of the ordination (Legendre and Legendre 2012). In the test dataset we visualized the variation of over 12,000 known and unknown root metabolites among populations using an NMDS with a continuous Jaccard distance. Permutations were undertaken to find a model with stress under 0.1.

If maintaining the distance between samples is preferred/required, often for use further in the study PCA is appropriate (Michalet et al. 2013, Hu et al. 2018, Wang et al. 2019). Unlike NMDS, PCA builds multiple independent axes explaining the variation in data, instead of compressing axes. Examining the first few axes can help visualize most of the variation in a

metabolomics sample, but not all of it, as each axis will contain independent information. Because the variation explained by each axis decreases sequentially, most of the variation in a metabolomics data matrix should be in the first few dimensions. Because PCA preserves the distance of the points, it is commonly used when those axes are intended for future use in models. Despite not currently seen in soil metabolomics studies, these PC axes could be used instead of specific compounds as fixed effects in models looking to examine the effect of metabolites on ecosystem processes. In the test dataset we used PCA to visualize the 12,000 distinct known and unknown root metabolites among populations. While the variation in PCA is split into more than two axes, we plotted the site scores of the first two axes to show as much variation confined in two dimensions.

Van Krevelen diagrams, originally developed to identify fuel hydrocarbons, are another data visualization tool unique to metabolomics that simplify columns of chemical data to atomic constituents before plotting samples in simplified space (Van Krevelen 1950). Van Krevelen diagrams compress chemical complexity into two axes, normally carbon to hydrogen and carbon to oxygen ratios, and an unknown compounds position in these two dimensions is indicative of its chemical family (Minor et al., 2014; Brockman et al., 2018, Mueller et al. *in review*). While estimates of carbon (C), hydrogen (H), and nitrogen (N) can be difficult to ascertain for unidentified metabolites, new software methods in identifying unknowns make it easier to estimate the atomic ratios of molecules in a sample (Kind & Fiehn, 2007; Tautenhahn et al., 2012; Kuhl et al., 2012). Comparisons using Van Krevelen diagrams are a distinctly qualitative approach, dismissing all information about relative abundances to instead focus on the chemical families of a sample, which can provide informative categories to help ascertain function, especially with unknown compounds that are shown to be important in the samples (i.e., using indicator analysis as seen in Mueller et al.*in review*). Comparisons among important compounds between samples can be undertaken without certain knowledge of the compounds themselves. In the test dataset we used Van Krevelen diagrams to explore the chemical families of the known and unknown root and soil compounds that separated populations. Indicator species analysis (Hill et al. 1975, De Cáceres et al. 2010) was used to determine the metabolites which were distinct in each population. In the case of this dataset, only unknown metabolites were identified with the analysis, making Van Krevelen diagrams useful for understanding the types of compounds that distinguished populations from one another.

Hypothesis testing in multivariate space

Much like a PCA, RDA creates independent perpendicular axes from a data matrix to visualize the multi-dimensional dataset in fewer dimensions. Similar to a regression however, it does so first by constraining axes by independent variables. This allows the researcher to explore the proportion of variation in the metabolite data matrix that can be explained by some external factor, such as population, location, or sampling time. An RDA can then plot these constrained axes to visualize how the independent variable shaped the relationship among samples in multivariate space. Redundancy analysis can be compared to a null model via permutation for the sake of hypothesis testing. Redundancy analysis is a tool currently underutilized by soil ecologists using metabolomics. In the example dataset we used the raw root metabolite data to assess if variation in the metabolites was explained by the sample's population, relative to a null model.

Similar to RDA, PLS-DA explores the variation in a metabolomics dataset by user-defined categories (i.e., factors of interest; Brereton & Lloyd, 2014). What makes PLS-DA different from RDA, is that the metabolome data is the x axis, and the PLS-DA attempts to fit metabolites into the categories defined as the y axis. With the example dataset, PLS-DA was used to bin root metabolites into a specific population or group of populations and then ascertain the importance of those populations compared to groupings of fewer bins. This analysis then permutes the strength of the assigned groups with leave-one-out analysis to determine how robust the categories are, resulting in an error rate (how often mis-categorization happens). PLS-DA, however, is not appropriate as the number of categories increases because the data matrix can only be manipulated once, despite the number of growing categories (Brereton and Lloyd 2014). As the number of distinct groups increases (or the researcher wants to account for a random variable or hierarchy) the PLS-DA error rate is likely to increase; therefore, this statistical method has specific uses when it is most appropriate. We use PLS-DA to test specific hypotheses within the data set from comparing two populations to six populations to show how its ability to distinguish differences among groups decreases as the number of groups increase.

If the hypothesis is instead testing if one data matrix (a dataset of metabolites) is related to another data matrix (another set of metabolites or perhaps genetic data of each sample), the Mantel test may be used. The Mantel test calculates the distance matrices of each data set and

then asks if distances associated with a sample are similar between the two matrices, calculating a correlation between them (Mantel 1967). With the test dataset, we calculated distances of both the root and soil metabolomes using a Jaccard distance before calculating the correlation between the root metabolite distance matrix and the soil metabolite distance matrix. Often however, the samples being compared are not independent variables, and there is some hierarchical structure (like population or site) to the data. It has been shown that performing partial-Mantel (Sokal 1979, Smouse et al. 1986) tests to account for this is often inappropriate (Raufaste & Rousset, 2001; Legendre & Legendre, 2012; Guillot & Rousset, 2013) and so we did not use this technique.

While the hypothesis testing of a Mantel test is only appropriate in two dimensional comparisons of distance with independent random variables, co-inertia keeps the data in multivariate space and explores a multivariate correlation (Dolédec & Chessel, 1994) which can overcome the limitations of Mantel, such as the sub-structure in the data such as population variation. Both tests can explore how correlated (between -1 and 1) the two matrices are, but co-inertia can draw that correlation for more than two axes. Using the example dataset, a multivariate correlation coefficient was calculated between the full root metabolome and full soil metabolome datasets using co-inertia. This analysis, in contrast to the Mantel test, finds the correlation in full multi-dimensional space before collapsing the visualization down to two axes, allowing more freedom in the substructure of the two data matrices.

Results

Ordination & Data visualization

NMDS

Non-metric multi-dimensional scaling of the root metabolite data successfully compressed the metabolite complexity to two axes (Figure 4.1a, stress < 0.1) showing the dissimilarity among populations. While this method can show the relationships among all of the data, NMDS does not maintain measurable dissimilarity between points, thus this approach is unable to accurately compare distances between clusters. Clustering of the points in this ordination suggests that there is structure in the data by populations (Figure 4.1a). Particularly,

the population coded in black, occupies unique space that was missed or decreased by PCA and PLS-DA.

PCA

The first two principal axes explain 66.2% of the variation in the root metabolomics dataset (Figure 4.1b). While this method does a poorer job of separating the samples based on population, variation can still be examined with the most deviate population (black symbols) showing the highest degree of separation. However, compared to the NMDS, there is less distinct clustering, and much greater overlap among populations. If only PCA were used to visualize this dataset, the conclusion that many of these populations have very similar metabolite communities would have been made erroneously. Having so much of the total variation explained on the first two axes make these good candidates for use in other models as predictors, potentially allowing the 12,000 known and unknown compounds of interest to be used as explanatory variables as only two parameters.

Van Krevelen diagrams

To visualize which chemical components of the dataset were responsible for the separation of populations in multivariate space seen above (Figure 4.1 a, b, c), Indicator species analysis was used to determine which unknown compounds should be represented on a Van Krevelen diagram (Figure 4.1d). Interestingly, it is only the unknown metabolites that are significantly driving the differences among populations, a result many past studies may have missed by analysis of the identified compounds only. The unknown compounds responsible for separating the population described with black symbols from Figures 4.1a and 4.1b can be seen here on Figure 4.1d. Despite incomplete knowledge of the chemical formula of these important compounds, these data suggest that this population's differences are from primarily saturated hydrocarbons, narrowing the scope for future identification.

Multivariate hypothesis testing

Partial least squares-Discriminant analysis

Despite the prevalence of PLS-DA in the broader metabolite literature (Gromski et al. 2015), the PLS-DA did not find sufficient separation between groups (Figure 4.2a). A misclassification rate of 0.5 suggests that this model is doing a poor job of fitting metabolites into categories based on population. An indicator species analysis may therefore be a better strategy, when combined with RDA for determining the metabolites that differentiate the populations or other groupings of interest. When used with fewer groups, as a subset or larger grouping of the data, as is more appropriate, the PLS-DA performs well. Distinct separation between groups can be seen for models including two through five populations (Figure 4.3a-d). The error rates for these tests remained under 0.25, half of the error rate found when all six populations are used (Figure 4.3e), suggesting that in this dataset, the PLS-DA is most useful in examining subsets of the data instead of the entirety, where RDA would perform better (Figure 4.3f).

Redundancy analysis

In support of the hypothesis that past population genetic divergence impacted plant root metabolomes, a redundancy analysis (RDA), where canonical axes were constrained by population, showed that plant population had a significant effect on root and soil metabolomes. The model including source population was better than the null model, explaining 29.8% of the variation in compounds found in plant roots ($p=0.032$, Figure 4.1c, Figure 4.2b). Graphing the first two constrained axes is also a useful visualization. Much like the NMDS above (Figure 4.1a) the RDA plot shows distinct separation of populations in the first two axes, allowing visualizing the pattern of interest compared to the unconstrained PCA, despite showing less of the total variation.

Mantel test

A Mantel test was unable to distinguish a correlation between the distance matrices associated with the root metabolites and the paired rhizosphere soils (mantel r = 0.014, p=0.472, Figure 4.2c). Based on this result, we would conclude that there is no relationship between the root metabolome and the surrounding soil metabolome. However, based on the redundancy analysis (Figure 4.2b), it is clear that there is hierarchy in the data based on population of origin. The variation among populations should be taken into account as this superstructure of the data will diminish the performance of Mantel tests which assume independence among samples. However, it is important to resist attempting to account for the effect of population with a partial Mantel test (Legendre and Legendre 2012, Guillot and Rousset 2013) as those tests often lead to incorrect type I error rates.

Co-inertia

Unlike a Mantel test, co-inertia analysis determined a relationship between paired root and soil metabolome data in a shared multivariate space. Contrary to the Mantel test, it found support for the hypothesis that root metabolomes influence soil microbial rhizosphere metabolomes. The significant ($p < 0.05$, via permutation) multivariate correlation coefficient (RV) of 0.385 suggests a good relationship between paired samples (Figure 4.2d). Many populations (arrow color) exist in unique areas of this shared multivariate space.

Discussion

Here we show, using the same dataset, that different visualization and statistical approaches have different strengths and shortcomings that can alter interpretation of metabolome results. Second, we show how utilization of whole metabolomics datasets, including unknown compounds, can provide unique insights, especially when visualized with Van Krevelen diagrams. Lastly, we outline the multiple multivariate statistical tools that can be applied to particular metabolomics questions to make appropriate analysis choices based on study goals. These approaches expand what has been typically used in animal and plant tissue metabolomics and should be applicable to a broad range of responses. These specific comparisons show how different conclusions would be made depending on which approach was used. We advocate for

particular approaches depending on specific questions addressed which may be unique to the analysis of environmental metabolomics hypotheses. We have shown how the complexity of the dataset can be an important factor for determining the most useful analysis.

Comparison of visualization techniques

While ordination techniques are not uncommon in metabolomics studies, the choice is often unstated or unclear. Both PCA and NMDS display the complex metabolomics dataset, but do so in different ways. As seen with the example dataset, both NMDS and RDA demonstrated the variation due to population better than the PCA. The PCA had the greatest overlap among points and only seemed to separate a few of the most deviate samples, while showing ~66% of the variation. Clearly, if visualization is the only goal, NMDS will show a more complete picture of the variation on only two axes and RDA will show variation explicit to some predicted factor. PCA has its place as a useful tool in soil metabolomics analysis if the axes of variation are used in subsequent analysis, allowing the chemical complexity of the soil to be modeled with only a few independent factors.

Soil samples are more likely to contain compounds that cannot be identified by the current libraries used in metabolomics, making the use of Van Krevelen diagrams highly recommended (Brockman et al. 2018). Instead of eliminating the majority of the data, as is typically done when unknown compounds are not used in analyses, we used Van Krevelen diagrams to visually demonstrate how important saturated hydrocarbons, lipids, and lignin were to the separation of root and soil metabolite communities among populations. Van Krevelen diagrams are being used in analytical chemistry, and their utility to identify broad classes of compounds continues to improve (Brockman et al. 2018). The root and soil metabolite datasets, like the one presented, contain a lot of information that when omitted undermines the exercise of untargeted analytical chemistry.

Comparison of hypothesis testing techniques

Despite the prevalence of redundancy analysis in genomic literature, it has yet to be fully embraced by soil ecologists studying metabolomics. This is surprising as the data collected in both genetic and metabolomic studies are very similar in structure (containing many more columns than samples, often reported in relative abundances), and often share the same goal of attempting to determine how function changes between samples across meaningful gradients. In

the case of ecological studies, the RDA is likely a better choice than PLS-DA when there are many groups, or multiple interacting effects that need to be examined for two reasons. First, PLS-DA is most informative in datasets where the majority of compounds have been identified, as it attempts to cluster metabolites into groups, which is not currently the case in metabolomics in complex environments. Secondly, it is easy to imagine situations where more than one predictor variable are used to explore the variation in metabolomes. While PLS-DA excels in binary categorization (i.e. treatment vs. control) and can be expanded to multiple categories (as seen in Figure 4.3), it is unable to explore multiple predictor variables as RDA is capable of comparing (McArdle and Anderson 2001, Gromski et al. 2015). Redundancy analysis is a strong tool for both visualization and hypothesis testing. As shown here, it resolved the complexity of a metabolomics matrix containing over 12,000 unique metabolites and determined the importance of plant population in this variation. As datasets become more complex, neither PLS-DA nor RDA may be appropriate. If a PLS-DA style approach is needed (i.e. identifying important distinct groups of compounds is the goal) but the data is too complex, random forest decision trees are a good alternative (see Cutler et al., 2007; Segal & Xiao, 2011) as they can handle many groups and hierarchical structure can be built into the decision tree. If the goal requires an RDA approach, but a third matrix is involved in the analysis (commonly genetic information or physical distance between sites), an analysis like PCNM should be used (Dray and Dufour 2007) which can compare populations to metabolites while accounting for distance or environmental factors. Many iterations of these analyses exist, and by having a clear understanding of the goals of a project, the appropriate decisions can be made for each question.

The pitfalls of relying on the Mantel test for analyzing the relationship between data matrices has been explored in detail multiple times both mathematically (Raufaste and Rousset 2001, Legendre and Legendre 2012) and with simulated data (Guillot and Rousset 2013). In this example dataset, the Mantel test was unable to find the relationship between paired root and rhizosphere soil metabolomes while co-inertia did show a relationship. This may be due to the hierarchy of populations present in the dataset. However, applying additional structure to a multivariate correlation with a paired Mantel test (Sokal 1979) to account for population can yield incorrect error rates based on the structure of the data and will often fail to address the intended biological hypothesis. If the samples are truly random variables, where each sample is independent, without an underlying structure in the experimental design, then the Mantel test can

be an easy to interpret analysis between two metabolomics matrices or to find a relationship between genomic and metabolomics datasets. Co-inertia, however, is a powerful tool for calculating the relationship between metabolomics matrices because it keeps the data in more than the two dimensions of the Mantel test while calculating the correlation. While more complex mathematically, co-inertia provides a correlation coefficient just as interpretable as the Mantel r but has the advantage of exploring the correlation in unreduced space, and accounting for the hierarchy of nested data (Dolédec & Chessel, 1994).

Conclusions and Recommendations

It is becoming easier and easier to obtain very large data sets associated with soil samples, from metabolomics, genomics, metagenomics, meta-transcriptomics as well as abiotic factors of sites from climate and geologic databases (Hampton et al. 2013) and despite the growth in data acquisition techniques, researchers often use common statistical methods. It is therefore critical to the advancement of soil ecology to embrace multivariate analysis, but often our inclination for analyzing these datasets returns to univariate methods or inappropriate partial-Mantel tests, despite their known problems (Raufaste and Rousset 2001, Guillot and Rousset 2013). While recent reviews call for Random Forests and other machine learning tools (Gromski et al. 2015), most metabolomics studies rely on a single visualization and analytic tool, with PLS-DA, PCA, and univariate linear models being the most common. As shown here, these approaches with this case study dataset have limitations that are common in ecological datasets, and would have failed to discriminate between the populations or to address the stated hypotheses with our test dataset. Building on these results and knowledge of the types of hypotheses that could be used in soil ecology, Table 4.1 is a guide for determining which test is most appropriate given possible hypotheses, data structure, and visualization goal. The table summarizes the range of common visualization and multi-variate hypothesis-testing approaches available for metabolomics datasets. It highlights the goal of each approach, the specific means to visualize the data and the specific type of data that must be used (i.e., type of matrix or category or if raw data can be used). The table also provides a range of hypotheses and the specific statistical test possible for each approach to aid in decision-making. Moreover, the table provides a referenced hypothetical example and the appropriate R package necessary to run each test.

This guide should enable best practices for researchers in soil ecology to determine which models best simplify the complexity of metabolome data to address specific research questions. However, as the analyses shown here with this case study demonstrate, scientists exploring environmental metabolomics should consider the strength of RDA for directional multivariate regression over PLS-DA, especially when there are more than two levels of factors to distinguish among, co-inertia for exploring the correlations between paired multivariate matrices over the Mantel test. Van Krevelen diagrams have high utility for analysis of the large number of unknown metabolites common to soil samples. If we wish to make the most of the complex data sets we are creating, we need to more carefully choose our visualizations and statistical hypothesis testing tools, getting more from the data we work hard to collect.

Acknowledgements

This paper would not have been possible without the many statistical discussions and debates with: James Fordyce, Stephanie Kivlin, Jessica Moore, Leigh Moorhead, Joe Bailey, Ian Ware, Tyson Paulson, Shannon Bayliss, Kendall Beals and Melissa Liotta, who also built the figures for this manuscript.

References

Brereton, R. G., and G. R. Lloyd. 2014. Partial least squares discriminant analysis: taking the magic away. Journal of Chemometrics 28:213–225.

Broadhurst, D. I., and D. B. Kell. 2007. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2:171–196.

Brockman, S. A., E. V. Roden, and A. D. Hegeman. 2018. Van Krevelen diagram visualization of high resolution-mass spectrometry metabolomics data with OpenVanKrevelen. Metabolomics 14:48.

De Cáceres, M., P. Legendre, and M. Moretti. 2010. Improving indicator species analysis by combining groups of sites. Oikos 119:1674–1684.

Caseys, C., G. Glauser, K. N. Stölting, C. Christe, B. R. Albrectsen, and C. Lexer. 2012. Effects of interspecific recombination on functional traits in trees revealed by metabolomics and genotyping-by-resequencing. Plant Ecology & Diversity 5:457–471.

Clasquin, M. F., E. Melamud, and J. D. Rabinowitz. 2012. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Current protocols in bioinformatics Chapter 14:Unit14.11.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

van Dam, N. M., and H. J. Bouwmeester. 2016. Metabolomics in the Rhizosphere: Tapping into Belowground Chemical Communication. Trends in Plant Science 21:256–265.

Dolédec, S., and D. Chessel. 1994. Co-inertia analysis: an alternative method for studying species–environment relationships. Freshwater Biology 31:277–294.

Dray, S., and A.-B. Dufour. 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. Journal of Statistical Software 22.

Eliasson, M., S. Rannar, and J. Trygg. 2011. From Data Processing to Multivariate Validation - Essential Steps in Extracting Interpretable Information from Metabolomics Data. Current Pharmaceutical Biotechnology 12:996–1004.

Gromski, P. S., H. Muhamadali, D. I. Ellis, Y. Xu, E. Correa, M. L. Turner, and R. Goodacre. 2015. A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. Analytica Chimica Acta 879:10–23.

Guillot, G., and F. Rousset. 2013. Dismantling the Mantel tests. Methods in Ecology and Evolution 4:336–344.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11:156–162.

Hao, M., H. Hu, Z. Liu, Q. Dong, K. Sun, Y. Feng, G. Li, and T. Ning. 2019. Shifts in microbial community and carbon sequestration in farmland soil under long-term conservation tillage and straw returning. Applied Soil Ecology 136:43–54.

Hill, M. O., R. G. H. Bunce, and M. W. Shaw. 1975. Indicator Species Analysis, A Divisive Polythetic Method of Classification, and its Application to a Survey of Native Pinewoods in

Scotland. The Journal of Ecology 63:597–613.

Holman, J. D., D. L. Tabb, and P. Mallick. 2014. Employing ProteoWizard to Convert Raw Mass Spectrometry Data. Pages 13.24.1-13.24.9 Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology 24:417–441.

Hu, L., C. A. M. Robert, S. Cadot, X. Zhang, M. Ye, B. Li, D. Manzo, N. Chervet, T. Steinger, M. G. A. van der Heijden, K. Schlaeppi, and M. Erb. 2018. Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. Nature Communications 9:2738.

Kind, T., and O. Fiehn. 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8:105.

Van Krevelen, D. 1950. Graphical statistical method for the study of structure and reaction processes of coal. Fuel 29:269–284.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27.

Kuhl, C., R. Tautenhahn, C. Böttcher, T. R. Larson, and S. Neumann. 2012. CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Analytical Chemistry 84:283–289.

Legendre, P., and L. Legendre. 2012. Numerical Ecology, 3rd english edition. Page (Developments in Environmental Modelling). Third edition. Elsevier Inc., Amsterdam.

Liland, K. H. 2011. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. TrAC Trends in Analytical Chemistry 30:827–841.

Mantel, N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Research 27:209–220.

McArdle, B. H., and M. J. Anderson. 2001. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. Ecology 82:290–297.

Mhlongo, M. I., L. A. Piater, N. E. Madala, N. Labuschagne, and I. A. Dubery. 2018. The Chemistry of Plant–Microbe Interactions in the Rhizosphere and the Potential for Metabolomics to Reveal Signaling Related to Defense Priming and Induced Systemic Resistance. Frontiers in Plant Science 9:112.

Michalet, S., J. Rohr, D. Warshan, C. Bardon, J. C. Roggy, A. M. Domenach, S. Czarnes, T. Pommier, B. Combourieu, N. Guillaumaud, F. Bellvert, G. Comte, and F. Poly. 2013. Phytochemical analysis of mature tree root exudates in situ and their role in shaping soil microbial communities in relation to tree N-acquisition strategy. Plant Physiology and Biochemistry 72:169–177.

Minor, E. C., M. M. Swenson, B. M. Mattson, and A. R. Oyler. 2014. Structural characterization of dissolved organic matter: a review of current techniques for isolation and analysis. Environ. Sci.: Processes Impacts 16:2064–2079.

Mueller, L. O., S. R. Borstein, E. D. Tague, S. P. Dearth, H. F. Castro, S. R. Campagna, J. K. Bailey, and J. A. Schweitzer. (*in review*). Populations of Populus angustifolia have evolved distinct metabolic profiles that influence their surrounding soil. Plant and Soil *in review*

Pearson, K. 1901. On lines and planes of closets fit to systems of points in the space. Philosophical Magazine.

Peters, K., K. Gorzolka, H. Bruelheide, and S. Neumann. 2018. Seasonal variation of secondary metabolites in nine different bryophytes. Ecology and Evolution 8:9105–9117.

Pétriacq, P., A. Williams, A. Cotton, A. E. McFarlane, S. A. Rolfe, and J. Ton. 2017. Metabolite profiling of non-sterile rhizosphere soil. Plant Journal 92:147–162.

Raufaste, N., and F. Rousset. 2001. Are partial Mantel tests adequate? Evolution 55:1703-1705

Ristok, C., Y. Poeschl, J.-H. Dudenhöffer, A. Ebeling, N. Eisenhauer, F. Vergara, C. Wagg, N. M. van Dam, and A. Weinhold. 2019. Plant species richness elicits changes in the metabolome of grassland species via soil biotic legacy. Journal of Ecology Early view.

Segal, M., and Y. Xiao. 2011. Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1:80–87.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. Systematic Zoology 35:627–632.

Sokal, R. R. 1979. Testing Statistical Significance of Geographic Variation Patterns. Source: Systematic Zoology 28:227–232.

Stough, J. M. A., S. P. Dearth, J. E. Denny, G. R. LeCleir, N. W. Schmidt, S. R. Campagna, and S. W. Wilhelm. 2016. Functional Characteristics of the Gut Microbiome in C57BL/6 Mice Differentially Susceptible to Plasmodium yoelii. Frontiers in Microbiology 7:1520.

Swenson, T. L., and T. R. Northen. 2019. Untargeted Soil Metabolomics Using Liquid Chromatography–Mass Spectrometry and Gas Chromatography–Mass Spectrometry. Pages 97–109 *in* E. Baidoo, editor. Microbial Metabolomics. Methods in Molecular Biology. Humana Press, New York.

Tautenhahn, R., G. J. Patti, D. Rinehart, and G. Siuzdak. 2012. XCMS online: A web-based platform to process untargeted metabolomic data. Analytical Chemistry 84:5035–5039.

Team, R. D. C., and R. R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Wakelin, S., S. Harrison, C. Mander, B. Dignam, S. Rasmussen, S. Monk, K. Fraser, and M. O'Callaghan. 2015. Impacts of endophyte infection of ryegrass on rhizosphere metabolome and microbial community. Crop and Pasture Science 66:1049.

Wang, Y., W. Ren, Y. Li, Y. Xu, Y. Teng, P. Christie, and Y. Luo. 2019. Nontargeted metabolomic analysis to unravel the impact of di (2-ethylhexyl) phthalate stress on root exudates of alfalfa (Medicago sativa). Science of The Total Environment 646:212–219.

Appendix

Table 4.1: A range of ordination and multi-variate statistical techniques for visualizing and analyzing metabolomics datasets. Presented also are goals of each test, how the data could be visualized, the specific requirements of the data and tests for particular hypotheses. A referenced hypothetical example is also shown, as well as R script resources required to run each analysis. The R packages are each available from the CRAN repository, and the necessary function is in the parentheses. The Seven Golden Rules is currently set up in Microsoft EXCEL and available from the Fiehn Lab (fiehnlab.ucdavis.edu/projects/seven-golden-rules).

Technique	Goal	Visualization	Data Requirements	Hypothesis	Test	Hypothetical Examples (and example citation)	R package
NMDS	Compress data matrix down to three or fewer axes in a way that preserves the relationships between points	Plots all of the variation of the metabolite dataset in visual dimensions, distance between points not metric.	A metabolite data matrix and a grouping vector (Site, Species, etc.)	Non-parametric cluster analysis is often paired with NMDS	Cluster analysis like ANOSIM are non-parametric tests of clustering among groups	Visualizing the separation of a metabolite data set based on groups like site or species [1]	Vegan(metaMDS)
PCA	Build axes that explain independent components of the variation in a data matrix	Axes that explain the most variation are plotted. Potential relationships in the data are hidden in un-shown axes	A metabolite data matrix	Independent axes explaining a proportion of the metabolite variation have predictive value	PC axes can be used as independent variables in a variety of statistical tests	Visualizing variation in a metabolite matrix[2]	Vegan(prcomp)

Table 4.1: Continued

Technique	Goal	Visualization	Data Requirements	Hypothesis	Test	Hypothetical Examples (and example citation)	R package
Van Krevelen	Visualize unknown metabolites using estimated atomic ratios	Ratios of H/C and O/C help visualize the likely chemical family of unknown metabolites	Atomic formula (at least estimated) of specific metabolites of interest	Variation among groups is explained by certain compounds	Important metabolites defined with PLS-DA, Indicator analysis. Univariate analysis of C/H and C/O ratios among groups possible	Once important, unknown metabolites are found, estimate molecular formula and use VK diagrams to determine which chemical families explain the variation among samples [3]	Seven Golden rules Mida7(VKplot) *
Redundancy analysis	Build axes that explain independent components of the variation in a data matrix based on a predictor matrix	Axes that explain most of the variation based on the predictor matrix are plotted	A metabolite data matrix and a matrix of predictors (Site, Species, etc.)	Axis constrained by population better predict metabolite variation than the null model.	P value based on permutations compared to null model	Address if groups like site, or species are explaining some of the variation in your data[4]	Vegan(rda)
PLS-DA	Fit metabolite matrix into discrete categories, find metabolites correlated to these categories	PC axes rotated to best capture distances based on groups	Categories to be predicted by a metabolite matrix	Categories based on population can be estimated via the metabolite matrix	Error rate based on leave-one-out permutation	Test if metabolite matrix can be broken into groups based on site, species, etc.[5]	DiscriMiner(plsDA)

Table 4.1: Continued

Technique	Goal	Visualization	Data Requirements	Hypothesis	Test	Hypothetical Examples (and example citation)	R package
Mantel Test	Find the correlation of paired distance matrices	Correlation of distances between paired samples	Distance matrices based on paired metabolite data (Or other multivariate matrix, commonly genomic data)	There is a correlation between the paired distance of the X and Y matrix	P value based on permutations compared to null model	Test the relationship between samples across paired metabolite datasets or between metabolite and metagenome sets[6]	Vegan(mantel)
Co-inertia	Find the correlation of paired matrices	Relationship of individuals paired sides in a shared multivariate space. Size and direction of arrows represents strength of similarities	Raw data matrices of paired metabolite data	There is a multivariate correlation between the paired metabolome data of the X and Y matrix	Multivariate correlation coefficient	Test how similar the relationships between samples are across paired metabolite datasets[3]	Ade4(coin)

Table refrences: [1]Wakelin et al., 2015, [2]Wang et al., 2019, [3]Mueller et al., *in review*, [4]Hu et al., 2018, [5]Ristok et al., 2019, [6]Caseys et al., 2012, [*]Borstein & Mueller, *in development*

Figure 4.1: Comparisons of visualization techniques available for metabolomics analysis. Populations are shown as separate colors in a, b, and c, where shaded ellipses represent 95% confidence intervals around each population centroid. Filled circles are each genotype's position in multivariate space. a) An NMDS compresses the relationship among individuals down to two axes. The population represented by black symbols stands out as separate from the others. b) The first two axes of a PCA show 66.2% of the variation in soil metabolites between samples. Here the populations overlap making it difficult to examine how these populations differ. c) RDA analysis with population defining the constrained axes were better than the null models, explaining 30% of the variation in compounds found in roots ($p=0.032$). The further two populations are from each other represent how different their metabolite communities are from each other. d) A Van Krevelen diagram showing atomic oxygen (O) to carbon (C) ratios along the x axis and atomic hydrogen (H) to carbon (C) ratios along the y axis of the indicator compounds which separate the most distant population (black symbols) from the other populations. Boxes estimating the range of chemical classes were sourced from Minor et al. 2014.

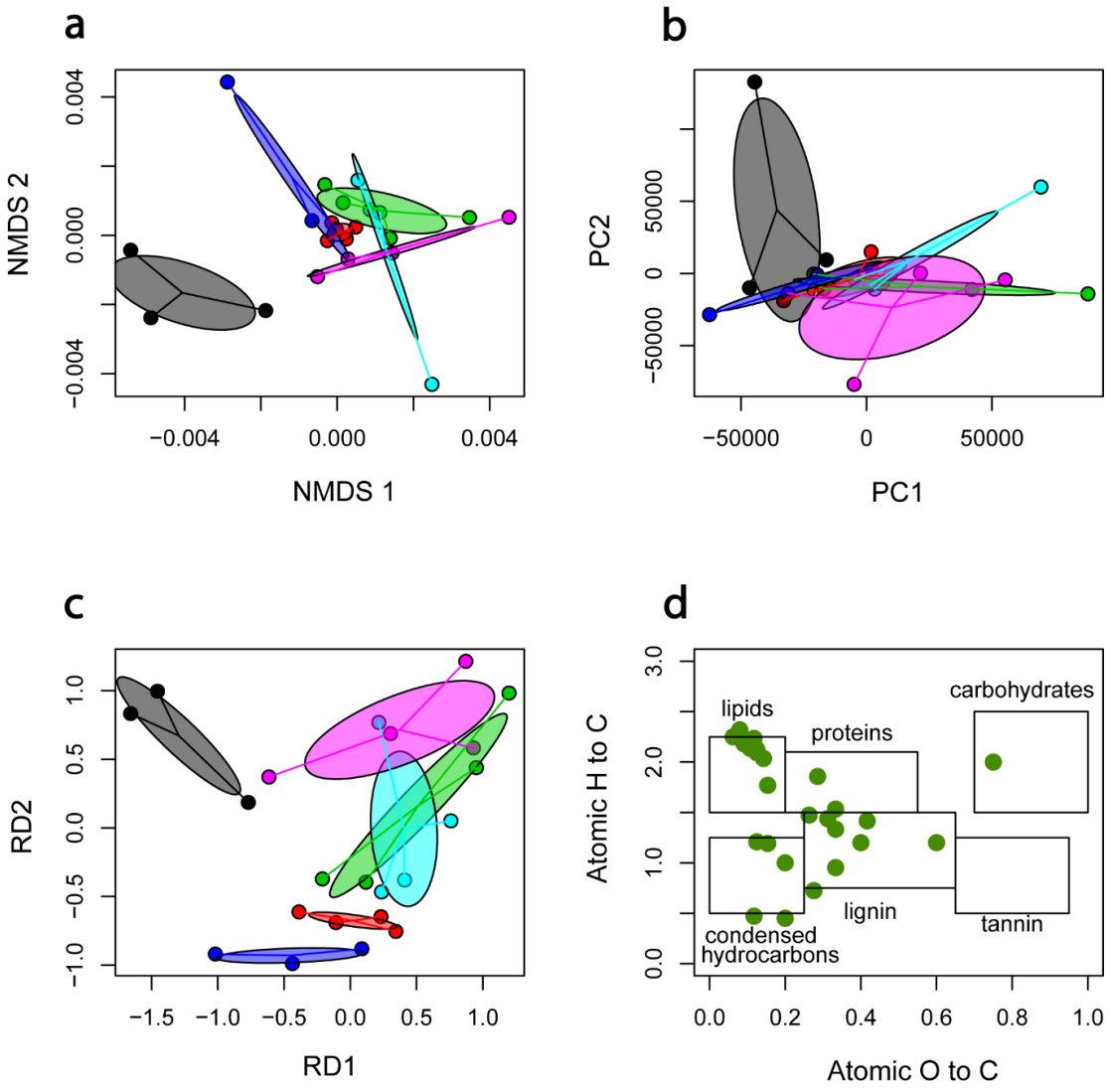

a
NMDS 2
0.004
0.000
−0.004
−0.004
0.000
0.004
NMDS 1
b
PC2
50000
0
−50000
−50000
0
50000
PC1
c
RD2
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
RD1
d
Atomic H to C
3.0
2.0
1.0
0.0
lipids
proteins
carbohydrates
lignin
tannin
condensed
hydrocarbons
0.0
0.2
0.4
0.6
0.8
1.0
Atomic O to C

Figure 4.2: Visualizing the ordinations associated with different hypothesis testing techniques. Colors represent the different plant populations. In panel a and b, ellipses represent 95% confidence intervals around each population centroid. Filled circles are each genotype's position in multivariate space. a) Visualization of the poor separation of a PLS-DA analysis across the six populations. A misclassification rate of 0.5 suggests that this model is not sufficiently separating each population, likely due to the number of groups being too large for classification. b) An RDA analysis is a better way to test the influence of many groups on a multivariate response. Redundancy analysis is not sensitive to the number of groups in the same way as PLS-DA. In this case RDA analysis with population defining the constrained axes were better than the null models, explaining 30% of the variation in compounds found in roots ($p=0.032$). Panels c and d show differences in multivariate correlations between multiple metabolite matrices. c) The relationship between root metabolite distance between samples on the x axis and soil metabolite distance on the y axis was slight (mantel $r = 0.014$) and not significantly different from zero ($p=0.472$) suggesting that there is no relationship between root distance and paired soil distance. d) Co-inertia axes 1 and 2 showing the relationship in shared multivariate space between root metabolites (base of arrow) and their paired soil metabolites (head of arrow) of each individual ($RV = 0.385$, $p < 0.05$). Distance between the head and tail is relative to the strength of the paired plant-soil relationship (smaller arrows are more tightly linked plant-soil signal). Distance between arrows is relative to the similarity between samples. Arrows closer together represent plants and soils that share similar root and soil metabolomes.

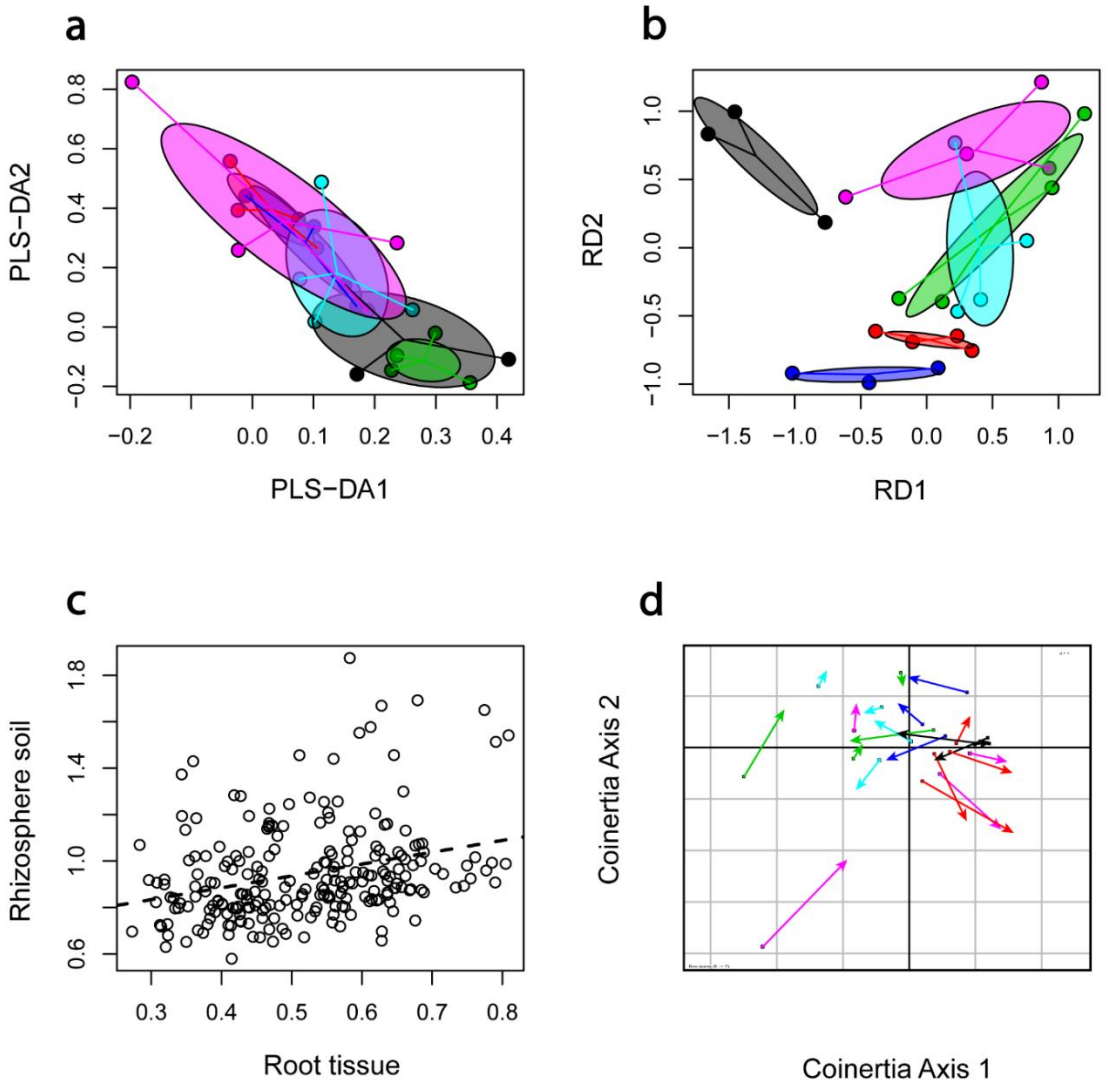
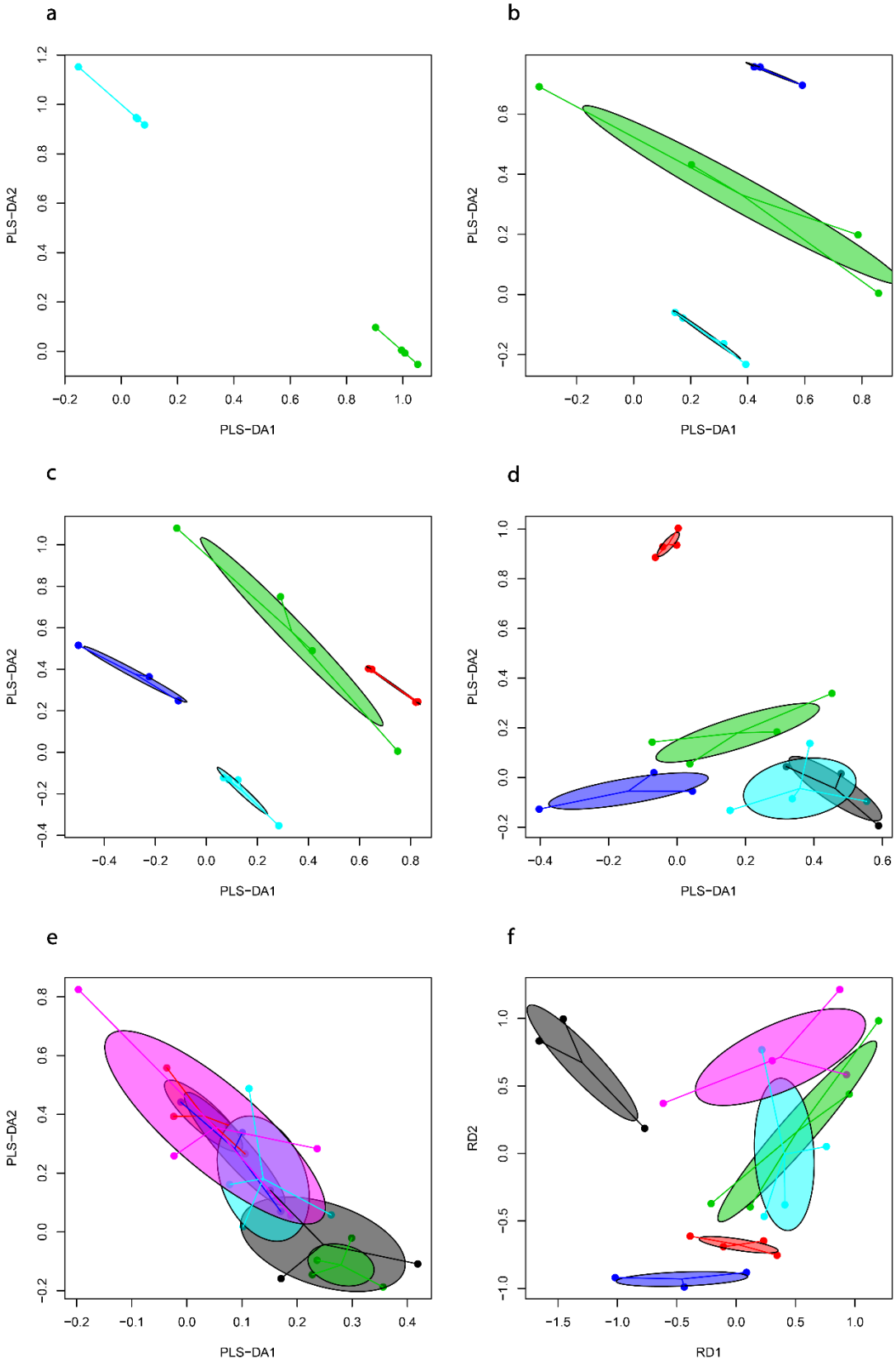

a
PLS−DA2
PLS−DA1
b
RD2
RD1
c
Rhizosphere soil
Root tissue
d
Coinertia Axis 2
Coinertia Axis 1

Figure 4.3: Visualizing the utility of PLS-DA with increasing groups. Visualization of the separation of a PLS-DA analysis from two to six populations. a) Two populations separate perfectly in space, with an error rate of 0.00. b, c, and d) Three, four, and five populations also separate well in this data set with a PLS-DA, with error rates of 0.18, 0.07, and 0.22 respectively. Once all six populations are in the model (e), the utility of the PLS-DA decreases and the model has an error rate of 50%. At this size, a redundancy analysis (f) often better represents the structure of the populations in multivariate space.


a
PLS-DA2
PLS-DA1
b
PLS-DA2
PLS-DA1
c
PLS-DA2
PLS-DA1
d
PLS-DA2
PLS-DA1
e
PLS-DA2
PLS-DA1
f
RD2
RD1

CONCLUSION

Our understanding of the distribution of terrestrial ecosystems on earth is quickly modifying with the introduction of new tools and the collaboration of multiple disciplines. It is insufficient to model the biosphere at the level of species, only interacting with large scale forces, and instead we must understand the mechanisms with which individuals interact with their surroundings. The field of eco-evolutionary biology aims to do this by discovering the feedbacks and consequences of trait evolution and niche construction. By working with foundation tree species in the field, common garden, and greenhouse my dissertation has expanded our understanding of the eco-evolutionary consequences of plant-soil interactions in two ways. First, across two different ecosystems I show how functional traits evolve in foundation tree species in response to abiotic and biotic, above- and belowground selective gradients. Second, I demonstrate the degree to which soil biotic communities can interact with plant genetic variation. The majority of which happens at the cellular scale, where novel use of metabolomics analytical and statistical techniques have shed new light on the diversity and complexity of the chemistry of the rhizosphere.

Chapter 1 shows how the evolution of plant traits occurred in short time scales (~150 years) and small spatial scales (kīpuka and matrix populations were separated by less than 200 m) a result not often seen in natural plant succession, but important to understand as fragmentation and disturbance become increasingly common. The divergence in ʻŌhiʻa, likely caused by the selective pressure of colonizing young lava substrates, altered traits which changed soil development, linking rapid evolution and ecosystem function. As global change puts more terrestrial ecosystems in similar scenarios, understanding how quickly trait evolution can change ecosystem function is an important finding. Furthermore, this chapter highlighted the importance of working in both the field and common garden, as demonstrating evolutionary variation in the controlled environment of a greenhouse and ecosystem change in the field simultaneously is a powerful tool for ecosystem science. Across larger spatial and temporal scales, in Chapters 2 and 3, using *Populus angustifolia* as a model system I demonstrated how variation in climate, soil, and surrounding biotic interactions have altered the population genetic variation in root metabolism. As I found in the tropics, the population level variation in plant traits was demonstrated to influence surrounding soil properties, conditioning unique soil metabolomes, even when the starting soil conditions are different. These studies offer avenues of exploration

for the chemical mechanisms that drive plant-soil interaction. By incorporating state-of-the-art analytical chemistry with modern multivariate statistics, I was able to demonstrate for the first time the enormous scale of the chemical complexity (tens of thousands of unique metabolites) which ties together the rhizosphere, while also showing the variability of that complexity across plant populations.

Using space for time substitutions, the soil inoculations in Chapter 3 show the relative importance of tree genotype to the root and rhizosphere metabolome as plants encounter novel environmental interactions (in this case new soil microbiomes). While future work is needed to elucidate the complete interaction between plant populations and soil microbiomes, this work represents an important early step in understanding how plant and soil traits interact to create distinct chemical communities in the rhizosphere and how these interactions may shape new, unpredicted ecosystem change.

By incorporating a paired root and soil sampling design into environmental metabolomics, Chapters 2 and 3 introduced new methods of analysis into the field. Examining both root tissues and surrounding rhizosphere soil allowed never before distinctions between metabolite sources. Using co-inertia, Chapter 2 was the first attempt at determining the metabolite effect of plant conditioning on rhizosphere soils, and Chapter 3 was the first attempt at examining the metabolome change in a factorial soil inoculation experiment. These experiments are the first to examine the entirety of a plants chemical conditioning of surrounding rhizosphere soils. Earlier attempts at using metabolomics in the understanding of plant roots often oversimplified the study system with hydroponics or sterile sand, oversimplified the sampling technique by examining a single component of the rhizosphere (i.e., root tissue and pot level exudates), oversimplified the system (i.e., not exploring genetic variation among samples), and oversimplified their analyses. Chapters 2, 3, and 4 aimed to remedy those oversimplifications by introducing novel experimental designs and analyses new to metabolomics (but common to ecology) in order to better understand the chemistry that underlies the relationships between plants and their soil surroundings. The interdisciplinary approaches used in this dissertation represent the way forward for predicting plant-soil interactions and the ecosystem level consequences of plant evolution.

VITA

Liam O'Connor Mueller was born to parents Charles W. Mueller and Maura E. O'Connor on the 19th of October 1990 in Honolulu, Hawaii. He attended elementary school at Waiʻalai Elementry, middle school at Assets School, and High School at Maryknoll High School. Liam attained a bachelor's of science in 2012 from The Evergreen State College in Olympia Washington. Here he was trained as an ecologist, analytical chemist and teacher. While in Olympia, Liam also worked for the United States Forest Service as a summer field technician, and for The Evergreen State College as an analytical instrumentation specialist. In 2013 Liam accepted a graduate teaching assistantship in Ecology and Evolutionary Biology from the University of Tennessee where he worked under Dr. Jennifer A. Schweitzer and graduated with a doctorate of philosophy in August 2019.